Therapeutic Targeting of the small RNA GcvB in *Actinobacillus pleuropneumoniae*

Daniela Lopes Cardoso

UNIVERSITY of PORTSMOUTH

The thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth

March 2021

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Daniela Lopes Cardoso

Date 10/03/2021

Word count: 44794

Acknowledgements

This thesis was largely written during one of the strangest years known to our generation due to the COVID-19 pandemic. It has been a time of extreme separation and isolation and also a time when contact with people has been more important than ever.

First and foremost, I want to thank my first supervisor, Professor Anastasia Callaghan, for the opportunity of embarking in this extraordinary adventure into and beyond the realms of science. Anastasia, you were the best supervisor I could ever have asked for. You are not only an expert in the field but a visionary leader in life. I have seen you overcome some extreme challenges and can honestly say that your contributions to my learning go beyond just science and firmly into the world of ethics and team cultivation. You are able to see the best in every person and support them develop their full potential. Thank you for all your help and support in getting me to this point, especially at the end of my PhD.

Another amazing human being I would like to thank is Dr Helen Vincent. Helen, I can see that you are not only invaluable to the team, but personally I find you to be one of the most altruistic people I have ever met. You have found time in your tight schedule to support me through to the end of my PhD. I will always remember your assistance at a really difficult time. Thank you.

To my second supervisor, Dr Colin Sharpe, it feels as though you have been in my scientific life since forever. Colin, as my first mentor you encouraged me to focus and persevere and believe that I could complete my masters and move into my PhD. You have been there during the final stages of writing my PhD too and helped to keep me calm and believe I could do it once again. Thank you.

My lab colleagues Layla, Callum, Bradley, Ciara, Martin and Dr Charlotte Henderson, I have missed your company whilst being away from the lab during the pandemic, but you have all been there when I needed you, as you always were when we could be in the same room. Layla, thank you for our great conversations and always being kind to me. Callum and Martin, I really enjoy our deep conversations, I always learn something new from them.

My former lab colleague, Dr Charlotte Caddick, who has now left the lab, thank you for all you did and for being so open to me when I first started at the lab, sharing information and helping me to find things in the freezer.

Dr Garry Scarlett and Dr Sam Robson, thank you for keeping my annual reviews interesting and challenging. I always dreaded them and came out having enjoyed the conversation. Thanks for asking so many questions and taking the time to extract the knowledge from me without me realising. Sam, you have provided so much insight into the data analysis side of my work. Having access to your knowledge and experience has been invaluable and I probably would have given up on some of the impossible to use software packages without having you to bounce ideas off and work things through with.

To our collaborators, Dr Janine Bossé and Professor Paul Langford. Thank you for the wealth of information which has underpinned significant aspects of my work. On a personal level, I have enjoyed the time we have spent working together, especially in Brazil. Janine, it has been such a pleasure to get to know you and to spend some time together working in Chichester. I hope to be able to continue working with you. Thank you.

To Sian and Kelvin, thank you for being such good friends and been there for me, throughout our undergraduate years to now. It all started in Brighton and I cannot believe it has been so long!

To my friends Wendy and Gemma, who were there when I needed a break and someone to listen to me when I most needed, as well as ad hoc childcare and technical trouble shooting. Thank you.

Mum, Gui, Renata and Renata (yes, my sister and sister-in-law are both called Renata, what are the chances?), thank you for knowing when I am stressed and that I don't mean it and for loving me anyway. Knowing you are there, even if I can only see you through a tiny FaceTime window has kept it real. Thank you, as always for being there.

To my darling Sophia, you have been amazing throughout these challenging times. I am extremely proud of you for working hard and becoming so independent lately, so I could complete my thesis. I am looking forward to spending more time with you and embarking in new and exciting adventures. Eu te amo, meu amor.

For last, but never least, I want to thank my wife and life partner, the other Mrs Lopes Cardoso. Fee, thank you for being there for me, especially now at the end of the Ph.D. You have been amazing, and it is true, I would not have made it without you. You are undoubtedly the best partner and the one to stick with. I am so grateful that our paths have crossed, and I am looking forward to our next chapter with our little family. I love you.

Research Dissemination

Manuscripts *in press*

1. **Generation of functional-RNA arrays by *in vitro* transcription and *in situ* RNA capture for the detection of RNA-RNA interactions**

Helen A. Vincent, Charlotte A. Henderson, **Daniela L. Cardoso** and Anastasia J. Callaghan

Methods in Molecular Biology (MiMB)

RNA performs a wide variety of vital cellular functions. These functions typically require interactions with other biological macromolecules, often as part of an intricate communication network. High-throughput techniques capable of analysing RNA-based interactions are therefore essential. Functional-RNA arrays address this need, providing the capability of performing hundreds of miniature assays in parallel. Here we describe a method to generate functional-RNA arrays using *in vitro* transcription of a DNA template array and *in situ* RNA capture. We also suggest how functional-RNA arrays could be applied to investigating RNA-RNA interactions.

Poster presentation

"**Developing sRNA-based therapeutics for disease caused by *A. pleuropneumoniae***"

The above title was used for multiple presentations throughout the course of this PhD:

1 – Presented by Daniela Lopes Cardoso at the Science Together conference, University of Portsmouth (2017).

2 – Presented by Daniela Lopes Cardoso at the Institute of Biological and Biomedical Sciences (IBBS) conference, University of Portsmouth (2018). Awarded first prize.

"**RNA Array for Exploring RNA Based Interactions**"

Presented by Daniela Lopes Cardoso at the South West Structural Biology Consortium Meeting (SWSBC), University of Reading, (2019).

Abstract

Actinobacillus pleuropneumoniae (APP) is a porcine respiratory Gram-negative pathogen, which has a profound negative impact on the farming economy, with serotype 8 being the most predominant in the UK. In anticipation of an antibacterial resistance era, there is increasing interest in the development of novel therapeutic strategies to combat this disease. Small non-coding RNAs (sRNAs), involved in post-transcriptional regulation of gene expression, are amongst some key components employed by bacteria to mediate virulence. sRNAs interact with their target mRNAs to tightly regulate the turnover of virulence factors, resulting in altered patterns of gene expression to rapidly adapt to environmental changes. Identifying, targeting and disrupting the production of virulence factors is a promising novel strategy to treat infectious diseases. Thus, sRNAs have been identified as potential antimicrobial targets. However, there are no therapeutic approaches targeting sRNAs being currently used at the moment.

This project aimed to understand and exploit the bacterial regulatory mechanism responsible for the control of virulence factor production by investigating and targeting regulatory sRNAs not yet characterised in APP. A combined approach of computational prediction and gene expression analysis using RNA sequencing data was designed, resulting in the establishment of a bioinformatic pipeline to identify novel *trans*-acting sRNAs from APP serotype 8 and predict their putative mRNA partners. Moreover, this research provides the first insights into the mechanisms of interaction of the sRNA GcvB and its predicted mRNA partners in APP. GcvB is a global regulatory sRNA involved in amino acid biosynthesis and peptide transport in various pathogenic Gram-negative bacteria. The expression levels of GcvB and its targets, quantified in different growth conditions, confirmed their presence in this species. The predicted interactions were then screened, validated and quantified using a novel RNA array technology developed at the University of Portsmouth with a customised protocol created to test APP molecules. After gaining a greater understanding of the mechanisms of sRNA-mRNA binding, a bespoke inhibitory nucleic acid mimic (NAM) was designed to specifically disrupt the interaction *in vitro*. The NAM molecule specifically disrupted the interaction between GcvB and its selected targets. Furthermore, the results showed that the novel high throughput RNA array technology can be applied and customised to identify and quantify interactions between sRNAs and their target transcripts. This work may have applications for novel therapeutics that target virulence pathways within the *Pasteurellaceae* family, and other bacterial species, that are mediated by sRNAs.

Contents

List of Figures

List of Tables

Abbreviations

A	adenine
A260	absorbance at 260 nanometres
aa	amino acid
AMP	Ampicillin
ANA	Anaerobic
APP	*Actinobacillus pleuropneumoniae*
APS	ammonium persulphate
ATP	adenosine triphosphate
AU	absorbance unit
bp	base pair
c	concentration
C	cytosine
cm	centimetre(s)
CTR	carboxyl terminal region
Da	Dalton(s)
dH2O	distilled water
DHfq	delta-Hfq
DMSO	dimethyl-sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotriphosphate
ds	double-stranded
DTT	dithiothreitol

E. coli	*Escherichia coli*
EDTA	ethylenediaminetetracetic acid
EMSA	electrophoretic mobility shift assay
EtBr	ethidium bromide
g	g-force
g	gram(s)
G	Gibbs free energy (deltaG)
G	guanine
GTP	guanosine triphosphate
HCl	hydrochloric acid
His-tag	6 histidine tag
HPLC	high pressure liquid chromatography
h	hour(s)
I	inosine
IPTG	isopropyl-β-D-thiogalactopyranoside
IVT	*in vitro* transcription
KCl	potassium chloride
K_d	dissociation constant
kDa	kilo-Dalton(s)
L	litre(s)
LB broth	Luria-Bertani broth
M	molar
mg	milligram
MG	malachite green
Mg^{2+}	magnesium ion

$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulphate
min	minute(s)
ml	millilitre(s)
mM	millimolar
mm	millimetre(s)
mol	mole(s)
mRNA	messenger ribonucleic acid
MW	molecular weight
NaCl	sodium chloride
ng	nanogram(s)
nL	nanolitre(s)
nm	nanometre(s)
nM	nanomolar
nt	nucleotide
NTD	amino terminal domain
NTP	nucleoside triphosphate
OD	optical density
OH	hydroxyl
ORF	open reading frame
P	phosphate
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PBST	phosphate-buffered saline with Tween 20
PCR	polymerase chain reaction

PDB	protein data bank
pg	picogram(s)
PS	porcine serum
RBS	ribosome binding site
RNA	ribonucleic acid
RNA-seq	ribonucleic acid sequencing
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
s	second(s)
SA	streptavidin
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulphate
sRNA	small ribonucleic acid
ss	single-stranded
T	thymine
T7	RNAP T7 RNA polymerase
TBE	Tris, Boric acid and EDTA
TBIO-PCR	thermodynamically balanced inside-out polymerase chain reaction
TEMED	N, N, N', N' - tetramethylethylenediamine
T_m	melting temperature
Tris	Tris (hydroxymethyl)-aminomethane
tRNA	transfer RNA
TY	tylosin
TXN	transcription buffer
U	uracil

UTP	uridine triphosphate
UTR	untranslated region
UV	ultraviolet
V	volts
v	volume
WT	wild-type
λ	wavelength
µg	microgram(s)
µl	microlitre(s)
µm	micrometre(s)
µM	micromolar
σ	sigma
$(NH_4)_2\ SO_4$	ammonium sulphate

Chapter 1

1. Introduction

1.1. Gene expression in bacteria

Bacteria are constantly exposed to changing environmental conditions, such as variation in the availability of nutrients or the presence of toxic agents that can cause cell distress. Thus, keeping proteome homeostasis is essential to maintain metabolic activity and guarantee cell survival. Production of proteins in prokaryotic cells is generally explained according to the central dogma of molecular biology, postulated by Francis Crick in 1958.

In this concept, all information necessary for life is stored in genes, which are stretches of deoxyribonucleic acid (DNA), consisting of a complementary double-stranded chain of nucleotides. The DNA is used as a template by the RNA polymerase (RNAP) enzyme to produce single-stranded ribonucleic acid (RNA) molecules in a process called transcription (Figure 1.1). The information stored in the RNA molecule codes for amino acid sequences, and macromolecular structures called ribosomes use this information to produce proteins in a process called translation.

However, a more elaborated mosaic of processes and pathways occurs around the central dogma to select which proteins are produced. These processes and pathways are regulated by tightly controlled mechanisms, collectively known as regulation of gene expression, occurring throughout transcription initiation to post-translational modifications. Under this premise, regulation of bacterial genes maintains a stable production of proteins by promoting the expression of genes involved in the relevant metabolic processes at a stage-specific pathway such as colonisation or evasion of host in pathogenic bacteria for example (Vitreschak *et al.*, 2004), and repressing the synthesis of gene products that are unnecessary at that specific time to conserve energy and nutrients (Balleza *et al.*, 2009; Hausser *et al.*, 2019).

Thus, regulation of gene expression in bacteria can efficiently occur at any time between transcription initiation and modifications for stabilisation of the final gene product. Various mechanisms of gene regulation in bacteria, of relevance to the work in this thesis, will be considered in more detail in later sections.

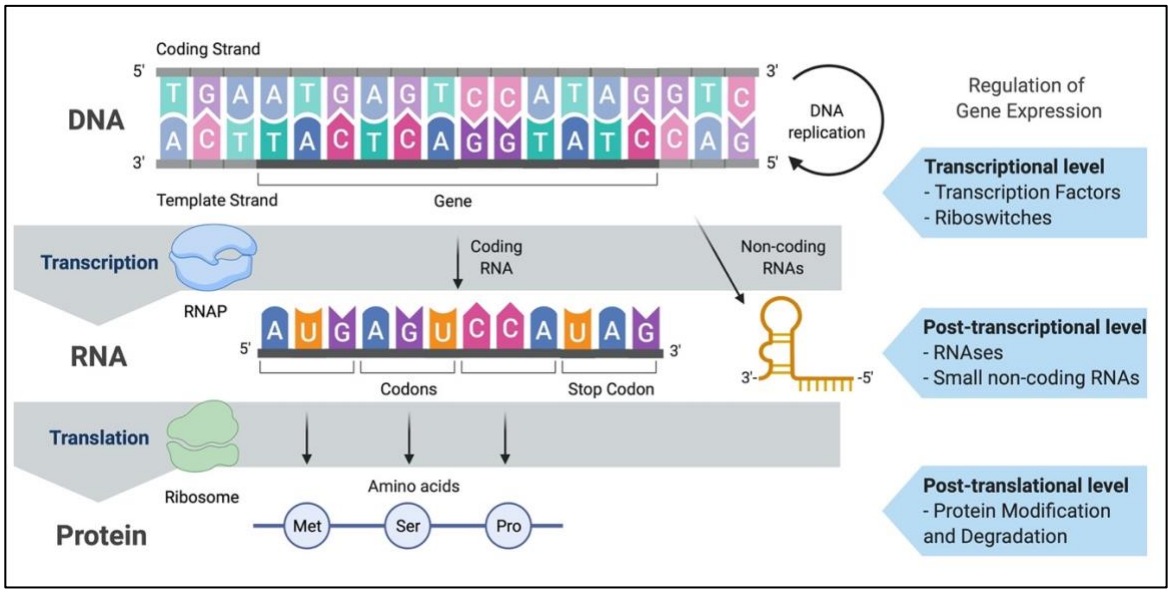


Figure 1.1. Schematic of Central Dogma by Francis Crick (1958). Deoxyribonucleic acid molecule (DNA) consists of two chains of nucleotides complimentarily connected by hydrogen bonds between their nitrogenous bases. Purine bases, guanine (G) and adenine (A), interact with the pyrimidine bases cytosine (C) and thymine (T), respectively. The double stranded DNA forms a helix structure. A delimited region of double stranded DNA containing a gene is used as a template (A is the template for uracil (U) rather than T) to synthesise ribonucleic acid (RNA) by the RNA polymerase (RNAP) in a process called **transcription**. During transcription initiation, RNAP binds to the DNA, separating the DNA strands to form a transcription bubble. Once transcription is terminated, if the newly made RNA molecule is a messenger RNA (mRNA), the sequence of its nucleotides is used to produce polypeptides and proteins by the ribosome in a process called **translation**. Three consecutive bases form a codon and each codon codes for a single amino acid. Gene expression may be regulated by a variety of mechanisms anytime in between these processes, and examples are noted in blue arrows. Some RNA molecules are not translated into a protein, acting as functional non-coding RNAs, such as small RNAs (sRNAs) in bacteria, which play a role in post-transcriptional gene regulation (PTGR). The mechanisms of gene regulation are further covered in this thesis.

1.1.1. Organisation of transcriptional elements in bacteria

Prokaryotic DNA is organised into a circular chromosome supercoiled in the nucleoid region in the cell cytoplasm, where transcription and translation may take place simultaneously (Stonington & Pettijohn, 1971). Genes can be transcribed individually as a mono-cistronic unit, or in groups as poly-cistronic mRNAs, organised in operons in the DNA sequence. A bacterial transcriptional unit can be defined by regulatory elements on the

DNA, including a promoter region where the RNAP binds, a transcriptional start site (TSS) followed by an untranslated region (UTR), one or more gene coding regions (ORFs) and a termination region (Figure 1.2). This format allows bacteria to respond quickly and efficiently to the signals they receive, since poly-cistronic mRNAs encode ORFs that generally act in the same metabolic route (Jacob & Monod, 1961; Browning & Busby, 2004).

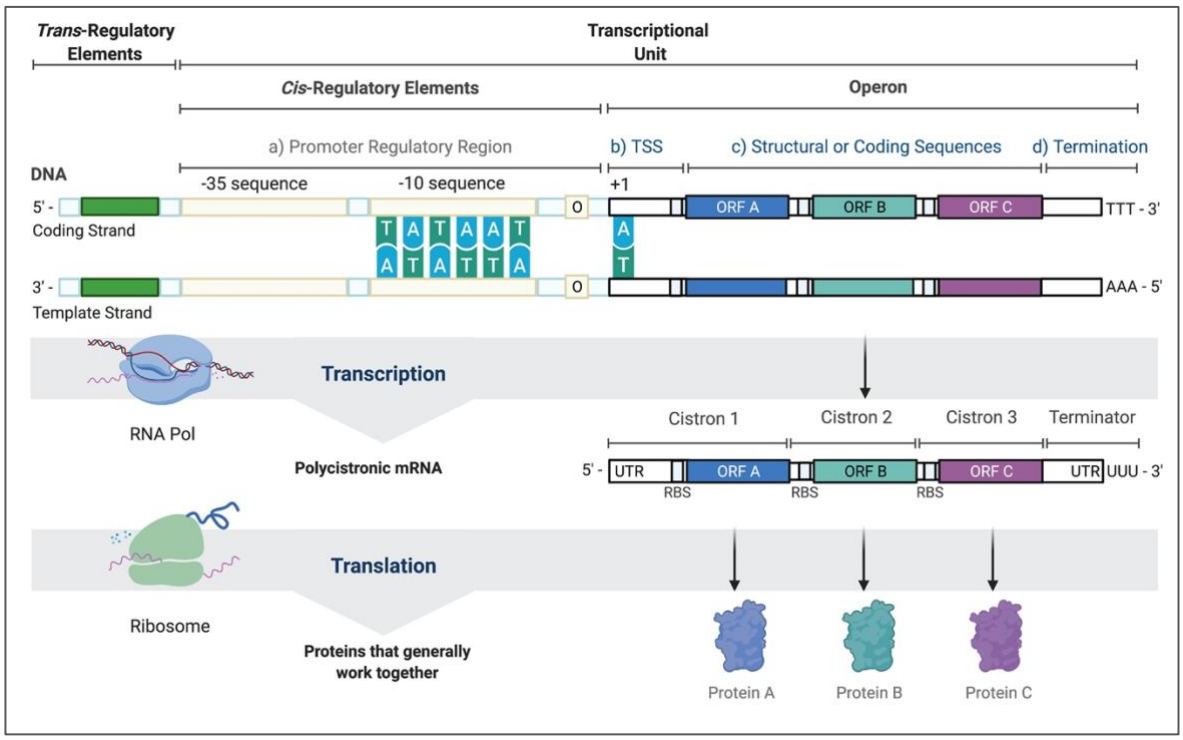


Figure 1.2. Typical prokaryotic transcriptional unit. The transcriptional unit which contains the **promoter regulatory region** (non-coding DNA sequences which regulate transcription of target genes) and the **operon** (containing the genes to be transcribed). **a)** Promoter regulatory region; containing *cis*-regulatory sequences recognised by the RNAP sigma (σ) factor protein; one at position -35 (consensus TTGACG) and another at position -10 (consensus TATAAT). This region may contain other regulatory sequences e.g., operators (O) on which repressor molecules bind, blocking the path for the RNAP thus disrupting transcription. **b)** Transcriptional start site (TSS) noted as position +1 in the unit. **c)** Structural coding sequences with ORFs encoding proteins that typically work together. **d)** Termination sequence containing signals to end transcription. The transcribed operon is a polycistronic mRNA containing more than one gene organised in cistrons, containing 5' UTRs adjacent to the coding region of each gene. The UTRs of a cistron may accommodate their ribosomal binding site (RBS) and other regulatory signals. Genes for ***trans*-regulatory elements** code for transcription factors, which are regulatory proteins that enhance or silence transcription and are usually located upstream of the transcriptional unit.

1.1.2. Transcription initiation

Transcription initiation is one of the most fundamental steps of gene regulation in prokaryotes. The start of transcription of a gene or operon depends on the RNAP

recognition of their promoters, and this is often assisted by additional mechanisms, involving the binding of molecules and regulatory proteins to specific regions of DNA called *cis*-regulatory elements, mainly located around the promoter of the target gene as illustrated in the schematics of the prokaryote transcriptional unit in Figure 1.2. This interaction may enhance or disrupt RNAP activity (Browning & Busby, 2004).

Before transcription initiation, the RNAP assembles as a holoenzyme involving two main components; an enzymatic RNAP core, which is responsible for the polymerisation of the ribonucleotides, and the sigma (σ) factor protein, necessary for the recognition of the transcription start site (TSS). The RNAP core has no specific affinity for the DNA, unless it is associated with a σ factor (Burgess & Travers, 1969; reviewed in Feklístov *et al.*, 2014). Transcription starts after identification of the promoter by the σ factor and the correct positioning of the RNAP holoenzyme on the TSS.

When the forming RNA chain reaches about 8-10 bases, the σ factor detaches from the RNAP core enzyme, which continues RNA synthesis on the DNA template strand until a termination signal is reached and the RNAP dissociated from the DNA, producing a complete RNA transcript (reviewed in Mooney *et al.*, 2005; Cases & De Lorenzo, 2005). Details of this process are illustrated in Figure 1.3. Regulation of gene expression at transcription initiation level is commonly performed by the RNAP σ factors, some of which are essential for successful infection in pathogenic bacteria, and consequently associated with virulence (Burgess *et al.*, 1969; Bashyam & Hasnain, 2004).

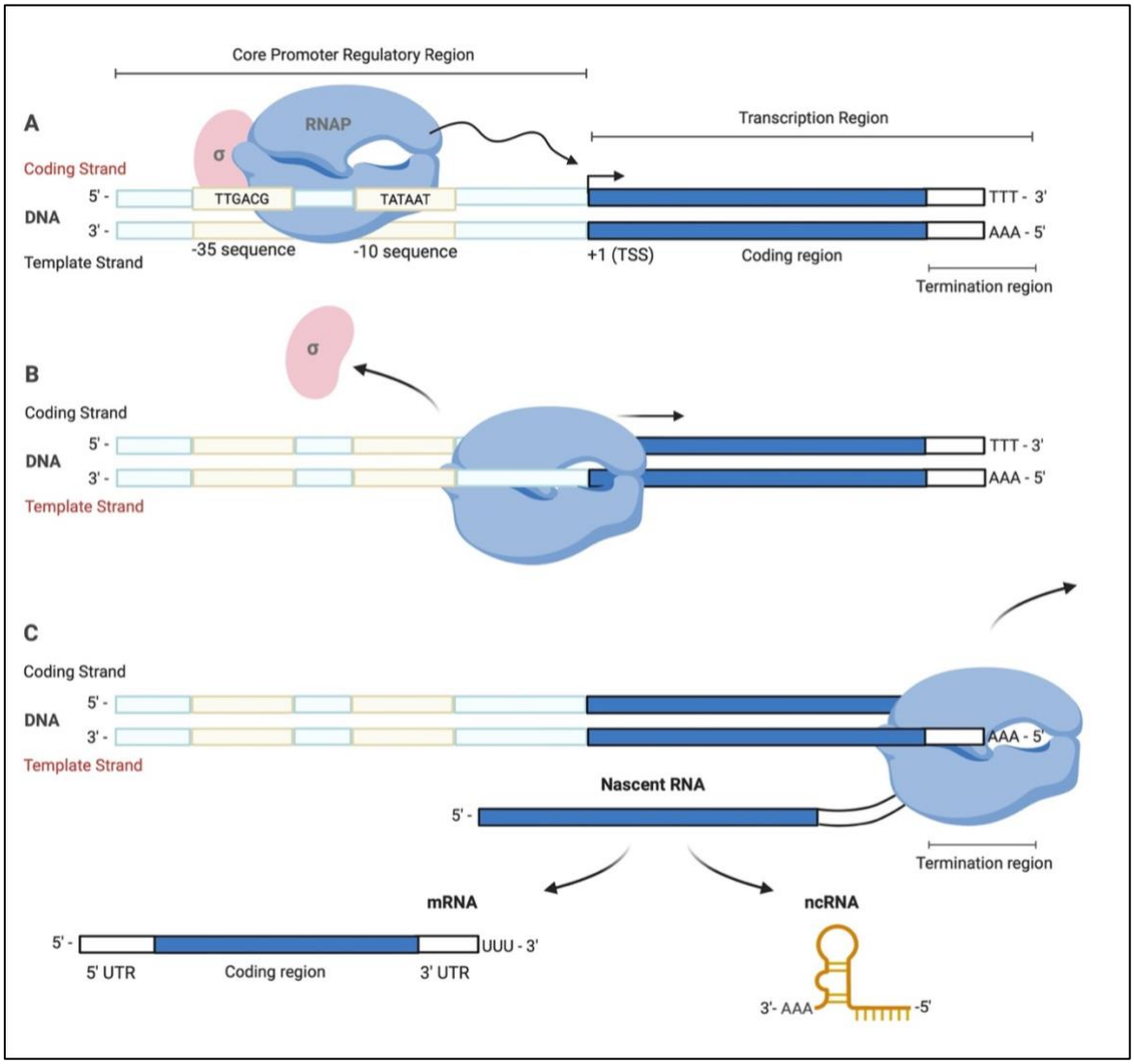


Figure 1.3. Transcription in bacteria. Diagram depicting generic transcription initiation in prokaryotes. **A)** Ahead of transcription initiation, the RNAP associates with a sigma (σ) factor which identifies and binds to promoter sequences at positions -35 (consensus TTGACG) and -10 (consensus TATAAT) on the DNA coding strand. This accurately places the RNAP in the direction of the target gene or genes. Once stabilised, the RNAP opens the DNA to start transcription on the template strand. The unwinding of DNA occurs at the -10 position, where the strands easily separate due to the predominance of the A-T base pairs that connect with only two hydrogen bonds, instead of the three from G-C base pairs. **B)** Once the sigma (σ) factor disassociates, the RNAP continues to transcribe throughout the ORFs until it finds a termination signal, and once this happens the RNAP transcribes a DNA sequence known as a terminator. **C)** The RNAP has finally reached the termination sequence, and a newly formed RNA may become a messenger RNA (mRNA), or a non-coding regulatory RNA (ncRNAs).

1.1.2.1. The role of transcriptional sigma (σ) factors

Other than recognising the promoter sequences during transcription initiation, bacterial σ factors play several key roles, including the opening and stabilization of the DNA complex in the promoter region; interaction with other regulatory molecules; and influencing the dissociation of RNAP from the promoter (Saecker *et al*., 2011; reviewed in Feklistov *et al*., 2014). Bacterial cells produce different σ factors, each responsible for regulating the transcription of a certain set of genes that handle conditions triggered by changes in the environment. Under a particular condition, a specific σ factor can direct the RNAP to the required genes to be transcribed. These external stimuli may include nutritional depletion, beginning of the stationary phase, temperature changes, or stress damage, such as oxidative and osmotic stress among others (reviewed in Ishihama, 2010).

Bacterial σ factors are divided into two families, the σ^{70} and the σ^{54} families, which are further divided in subgroups based on their phylogenetic origins (Gruber & Gross, 2003; reviewed in Murakami & Darst, 2003). This therefore creates competition between the different σ factors for the RNAP core, guided by the current condition that the bacterial cell is enduring (Aertsen & Michiels, 2004). All bacteria present at least one main essential σ factor, which is known as the housekeeping σ factor, and is responsible for recognising most bacterial promoters. The σ^{70} family comprises a greater number of factors related to the most varied functions during bacterial growth.

The members of this family are phylogenetically and structurally classified into four distinct subgroups, which are differentiated by the presence or absence of four conserved regions that correspond to four structural domains. These domains are associated with DNA binding in the -10 and extended -10 and -35 promoter regions (Gruber *et al*., 2001). Regulation by the RNAP transcription factors is highly dependent on DNA *cis*-regulatory elements associated to the target gene. A single gene with more than one promoter; can thus be regulated by more than one σ factor, which may be expressed in more than one condition that the bacterium cell is exposed to (reviewed in Feklistov *et al.*, 2014).

1.1.2.2. Transcription factors

Transcription factors (TFs) are regulatory proteins that recognize *cis*-regulatory elements in the DNA and promote the activation or inhibition of transcription of their target genes. TFs that stimulate transcription are called activators and act by stabilizing the binding of the RNAP to the promoter sequence. Their binding sites are usually located upstream of the -35 site but can also be found between the -35 and -10 sites. Meanwhile, TFs that prevent RNAP from accessing the promoter region are called repressors and their interaction sites are the operators usually found on the promoter region, e.g., regulation of the *lac* operon (Figure 1.4) (Jacob & Monod 1961; reviewed in Balleza *et al.*, 2009). Transcriptional *cis*-regulatory elements may overlap within operons and there is often a promoter that controls the transcription of the complete set of genes in the initial cistron, but there may be internal promoters regulating certain genes in the operon, or even premature transcription termination regions, causing only a few genes to be part of the poly-cistronic mRNA (Browning & Busby, 2004).

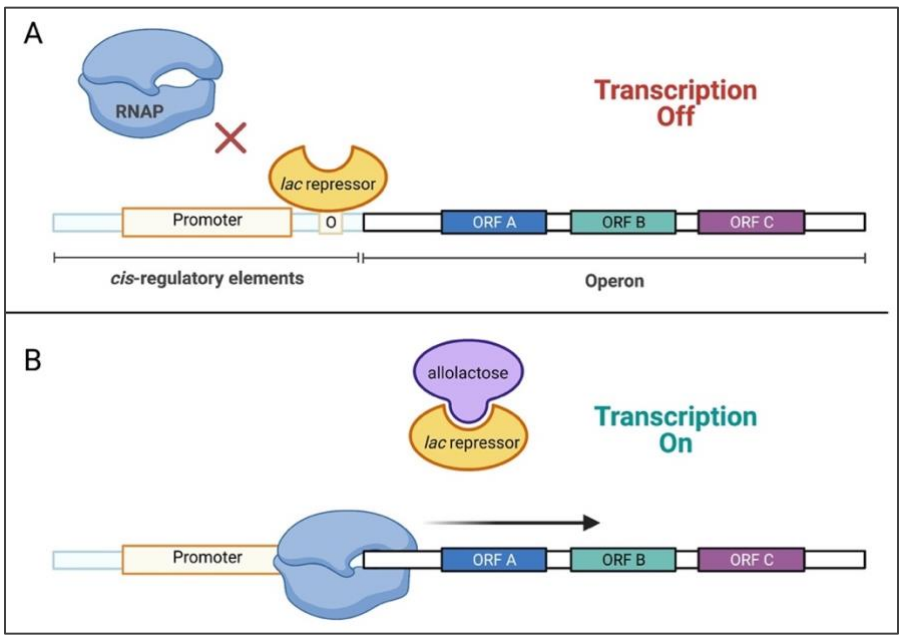


Figure 1.4. Regulation of the *lac* operon. The operon model was proposed as a mechanism for coordinated regulation of the gene expression related to lactose transport and catabolism by Jacob & Monod (1961). In this model, **A)** a repressor protein (*lac* repressor) binds on the operator, a *cis*-regulatory element sequence around the promoter region to which the repressor binds and controls the expression of the operon genes by blocking RNAP access to the promoter. **B)** In the presence of an inducer, in this case the allolactose protein, the *lac* repressor undergoes conformational change and loses affinity for the operator, resulting in de-repression of the operon and access by the RNAP resulting in the initiation of transcription (Lewis & Adhya, 2004).

1.1.3. Transcription termination

In prokaryotes, transcription terminators are divided in Rho-dependent and independent terminators. In the Rho-dependent termination model, the Rho protein binds to the mRNA at specific sequences during transcription and translocate downstream along the forming transcript towards the transcriptional machinery. The complex is then paused at the termination region, where the Rho protein contributes to releasing the RNA transcript by dislocating the RNAP (Matsumoto *et al.*, 1986; reviewed in Roberts, 2019). Alternatively, Rho-independent terminators usually contain palindromic sequences in the 3' extremity of the nascent RNA transcript, forming a stable stem-loop, followed by a conserved sequence of uracil residues known as 3' Poly-U tail. The formation of this termination structure in the in RNA disrupts associations between RNA and the RNAP and facilitate dissociation of the transcript (reviewed in von Hippel, 1998; Roberts, 2019).

1.1.4. Translation

An overview of translation in bacteria, such as *Escherichia coli*, is shown in Figure 1.5. During translation initiation, the smaller 30S subunit of the ribosome recognises the ribosomal binding site (RBS) and attaches to the 5' UTR, before moving along the mRNA molecule. The RBS region has consensus sequence 5'-GGAGG- 3' known as the Shine-Dalgarno sequence (Shine & Dalgarno, 1975). When the ribosome finds the start codon, the larger subunit (50S) binds to the smaller subunit (30S), and translation begins with the methionine amino acid AUG, and ends at the stop codon, usually UAG, UAA or UGA. As previously discussed, bacterial genes are often organised in groups and transcribed as polycistronic mRNAs. Each cistron unit may contain its own RBS and its gene ORF may be translated separately.

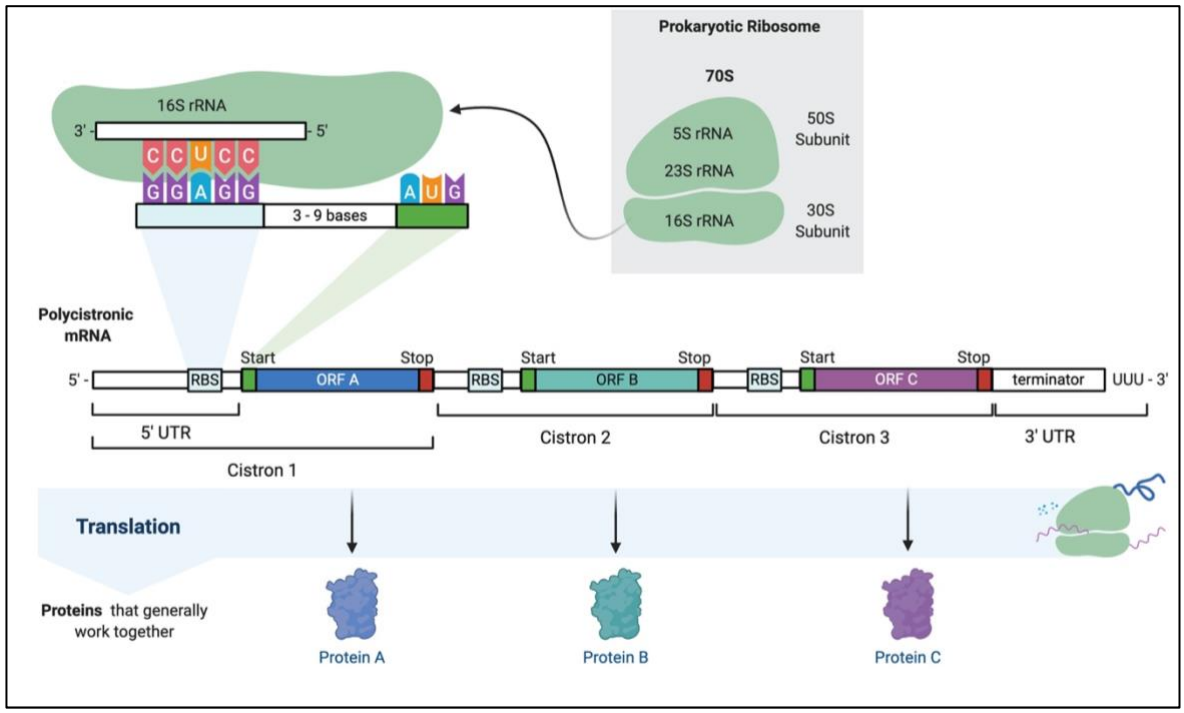


Figure 1.5. Translation in bacteria. The 70S ribosome consists of two subunits (30S and a 50S). The 30S subunit contains the ribosomal RNA (rRNA) molecule 16S and 21 different proteins, while the 50S subunit contains two rRNAs, 23S and 5S, and 31 different proteins. Key activities of translation such as mRNA decoding and incorporation of amino acids are performed by the rRNAs, instead of its component proteins. During translation initiation, the smaller subunit (30S) of the ribosome recognises the ribosomal binding site (RBS) and attaches to the 5' UTR, before moving along the mRNA molecule. The RBS region has consensus sequence 5'-GGAGG- 3' known as the Shine-Dalgarno sequence (Shine & Dalgarno, 1975). When the ribosome finds the start codon, the larger subunit (50S) binds to the smaller subunit (30S), and translation begins with the methionine amino acid AUG, and ends at the stop codon, usually UAG, UAA or UGA.

1.1.5. Regulation of gene expression in bacteria

Gene expression regulation in bacteria can occur at the transcriptional, post-transcriptional or post-translation level, as noted in Figure 1.1. However, having provided a general background introduction into the key processes of gene expression in bacteria, only regulatory mechanisms of relevance to the research in this thesis will be considered. Specifically, this will include post-transcriptional mechanisms - with a focus on those that control RBS access, sometimes with an associated impact on transcript stability, as a means of controlling protein translation.

1.1.5.1. Post-transcriptional gene regulation

Post-transcriptional gene regulation (PTGR) mechanisms, such as the regulation of translation, sometimes with associated impact on transcript stability, are essential in regulating metabolic pathways in prokaryotes, especially pathogenic bacteria (Browning & Busby, 2004). PTGR mechanisms can quickly act on the transcript levels, affecting its gene products as a quick response to environmental changes. This category of gene regulation includes the ability of the ribosome to access and bind the mRNA to initiate translation. Regulation of gene expression in bacteria can be mediated at a post-transcriptional level by translational riboswitches and regulatory small non-coding RNAs operating as *trans* (*trans*-acting sRNAs) or *cis*-acting antisense RNAs (asRNAs) (Winkler & Breaker, 2003; Mandin *et al*., 2013).

1.1.5.1.1. Riboswitches

At the post-transcriptional level, structures of the mRNA may form at important sequences of the mRNA involved in translation, such as the ribosome binding site (RBS) or at the translation start codon (reviewed in Mandin & Guillier, 2013). Regulation of gene expression by riboswitches is based on the perception of several intra and extracellular signals, such as presence of a particular molecule or variation in its concentration (Figure 1.6). For example, a lysine-responsive riboswitch in the mRNA of the *lysC* genes in *Bacillus subtilis*, changes conformation upon a concentration of lysine, resulting in the downregulation of the aspartokinase LysC, involved in lysine synthesis (Grundy *et al*., 2003; reviewed in Rodionov *et al*., 2003).

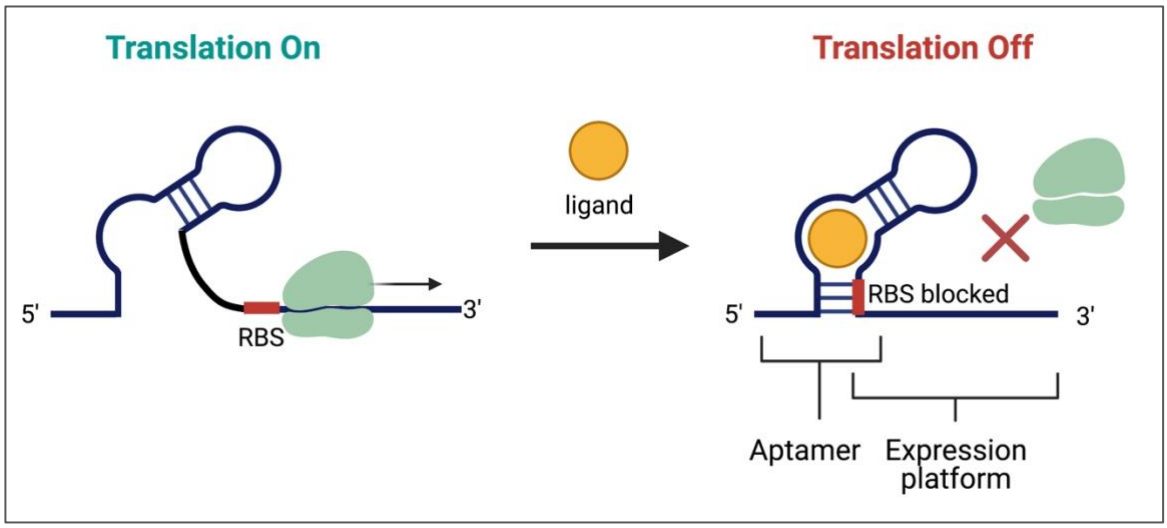


Figure 1.6. Generic diagram of post-transcriptional regulation by riboswitches. In this schematic, in the translational on state, the ribosomal binding site (RBS in red) is exposed for recognition and binding of the ribosome 30S subunit to initiate translation of the mRNA. In the presence of a ligand, the mRNA binding region (Aptamer) undergoes conformational change, sequestering the RBS sequence and inhibiting ribosome binding, and consequently translation initiation, affecting protein synthesis (adapted from Kim & Breaker, 2008; reviewed in Mauger *et al*., 2013).

1.1.5.1.2. Small non-coding RNAs

A class of molecules involved in post-transcriptional regulation of gene expression, the small non-coding RNAs (sRNAs), are amongst some of the key components employed by pathogenic bacteria to mediate virulence by regulating gene expression (reviewed by Vogel & Luisi, 2011). These molecules play an important role in several cellular processes, including sugar metabolism, transcriptional regulation, chromosomal replication, RNA processing and modification, protein inhibition, iron homeostasis and regulation of virulence in pathogenic bacteria (Man *et al*., 2011; reviewed in Svensson & Sharma, 2016). Currently, research on sRNAs is limited and represents a small proportion of the total RNA population in only a handful of selected bacterial species, although a study by Holmqvist *et al.* (2012) suggests that around half of *E. coli* genes may be regulated by sRNAs at some point in the bacterium cell cycle. As the focus of this thesis, more detailed insight into post-transcriptional gene regulation by sRNAs is covered in the next section.

1.2. Bacterial small non-coding RNAs

Bacterial small non-coding RNAs (sRNAs) are functional transcripts, usually between 50 and 500 nucleotides, produced in bacteria in response to environmental stresses, and have been observed in a multitude of species, including the widely studied *E. coli*. sRNAs function by forming a duplex with their target mRNAs through antisense base pairing on or around the ribosome binding site (RBS), causing diverse regulatory outcomes in protein synthesis (reviewed in Vogel & Luisi, 2011). They generally act as regulators of translation by altering access to the RBS to inhibit or activate translation. In situations of translation inhibition, there is often an associated impact on the stability of the target mRNAs targeting them for degradation (reviewed in Storz *et al.*, 2011; Vogel & Luisi, 2011; Mandin & Guillier, 2013).

Some sRNAs require a facilitator protein to perform their functions. The most studied and widely known is the chaperone Hfq protein, long known as the host factor for Q-beta RNA bacteriophage replication in *E. coli,* which appears to facilitate the action of many sRNAs on their target mRNAs, mainly to stabilize the interaction between these molecules (reviewed in Moller, 2002a; Vogel & Luisi, 2011). Hfq participates in regulation via sRNAs by forming the mRNA-sRNA-Hfq ternary complex. A brief schematic of the Hfq is illustrated in Figure 1.7.

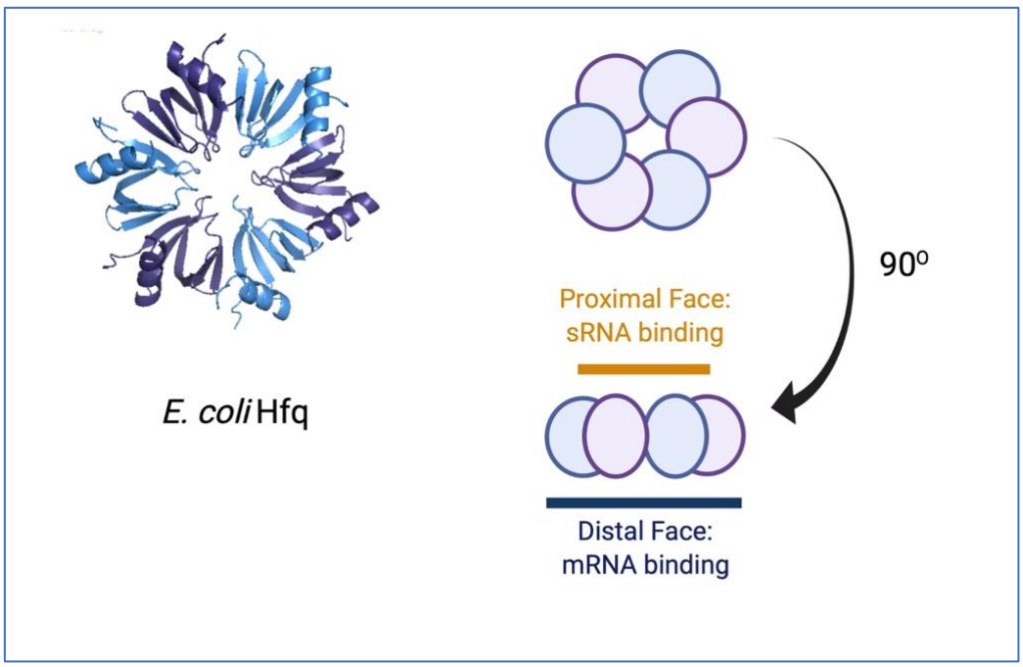


Figure 1.7. Components of the ternary complex sRNA-mRNA-Hfq. Cartoon depiction of the structure of *E. coli* Hfq (PDB accession code 1HK9) rendered in PyMOL. Hfq has a ring-shaped structure forming a homo-hexamer (subunits alternately coloured in blue and purple), with a proximal face (yellow line) involved in sRNA binding and stabilisation, and a distal face (blue line) that recognises binding motifs important for mRNA binding (Franze de Fernandez *et al*., 1972; Vogel & Sharma, 2005; Vogel & Luisi, 2011), although some mRNAs bind to both faces (Fender *et al*., 2010).

The first report on determining the function of sRNAs was the discovery of an RNA in *E. coli* complementary to the 5' untranslated region (UTR) of the *ompF* mRNA, coding for an outer membrane protein, resulting in the inhibition of its translation (Mizuno *et al*., 1984). Larger genomes require greater complexity in gene regulation, thus the number of sRNAs may vary with the size of the genome. The *E. coli* genome contains approximately 4,000 genes and 50 to 100 predicted sRNAs (reviewed in Gottesman, 2004).

Small RNAs may be encoded in the DNA strand opposite to that encoding their target mRNA (*cis* encoded), or at loci physically separated or distant from the open reading frame (ORF) that encodes their target mRNA (*trans*-encoded). *Cis* encoded antisense sRNAs (asRNAs) act on the respective genes on the opposite strand from the region from which they were transcribed, presenting perfect complementarity to their targets, generating stable interactions (reviewed in Thomason & Storz, 2010; Svensson & Sharma, 2016) as illustrated in Figure 1.8-A. There are reports of the effects of gene regulation mediated by asRNAs relevant to bacterial virulence and pathogenicity (reviewed in Mandin & Guillier, 2013; Svensson & Sharma, 2016).

Contrastingly, *trans*-acting sRNAs act on mRNAs encoded at distant loci in the genome from its targets (Figure 1.8-B). They interact with their target mRNAs via short partial base pairing to attenuate or activate their translation, and their secondary structure play a crucial role in transcript stabilisation and regulation (Wilson & Von Hippel, 1995; reviewed in Vogel & Luisi, 2011). *Trans*-acting sRNAs are versatile regulators due to their interaction by imperfect complementarity with the target, which consist of discontinuous base-pair matches forming structural bulges, and guanine-uracil (GU) wobble base-pairs, contravening the Watson-Crick base-pair rules (Crick, 1966). This allows for interaction between one sRNA and several mRNA targets (reviewed in Vogel & Luisi, 2011). Although asRNAs provide a fast regulation response by acting close to their transcription location, they only interact with targets encoded on the same genetic locus, while *trans*-acting sRNAs regulate a higher number of targets, allowing bacteria to adjust to various environmental stress signals at the same time (reviewed in Repoila & Darfeuille, 2009).

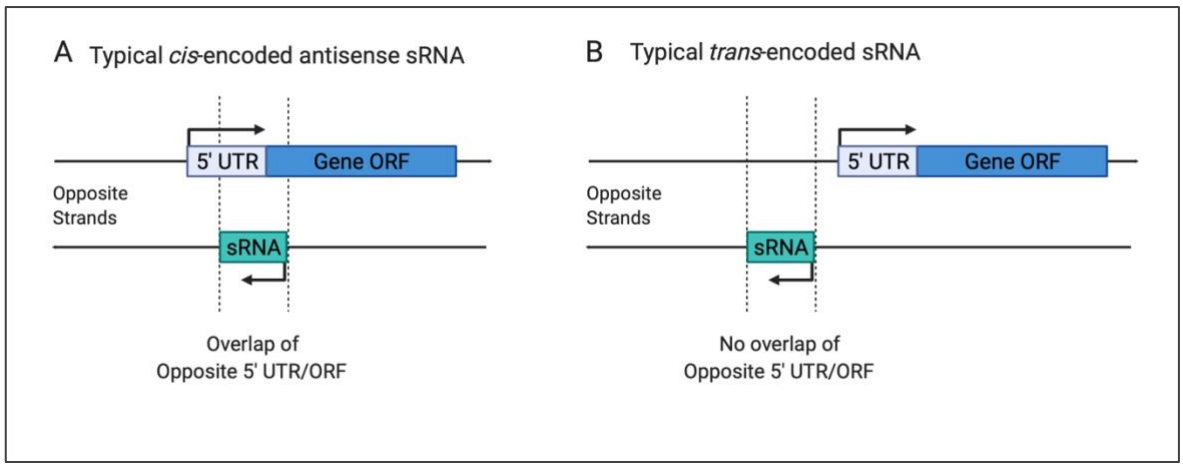


Figure 1.8. Diagram of *cis* and *trans*-encoded sRNAs. A) Genes for *cis*-encoded antisense sRNAs are usually located in the opposite strand of their target gene, overlapping its 5' UTR and ORF. **B)** *Trans*-encoded sRNA genes are in a different location from their targets and usually there is no overlap on genes on the opposite strand, and transcription signals such as transcription start site and 3' Poly-U are separated from neighbouring genes (Li *et al.*,2012).

1.2.1. Functional features of *trans*-acting sRNAs

Trans-acting RNAs are highly structured transcripts usually transcribed from intergenic regions (IGRs). They are generally mono-cistronic, having their own promoter and intrinsic transcription terminator. Most transcribed *trans*-acting sRNAs have a distinct secondary

structure containing stem-loops that facilitate interactions with other molecules. They present a functional interaction region, also known as seed region, that generally acts by short imperfect base pairing within a discrete region (usually a minimum of 7 bases) of their target mRNAs (Storz *et al.*, 2004; Vogel & Wagner, 2007; Sharma *et al.*, 2007; Vogel & Luisi, 2011). Seed regions have revealed a G/U-rich character in previous studies of sRNAs (reviewed in Updegrove *et al.*, 2015), including *Salmonella* GcvB (Sharma *et al.*, 2007), that target C/A-rich regions in the sequence of their mRNA targets. A poly-uracil tail (Poly-U) is a sequence of uracil residues that follows a stable stem-loop at the 3' end, considered an important regulatory signal for transcription termination (Wilson & Von Hippel, 1995). Many *trans*-encoded sRNAs, such as DicF, OxyS, Spot42 and GcvB in *E. coli*, silence their target mRNAs using mechanisms facilitated by Hfq (Bouché, 1989; Altuvia *et al.*, 1998; Urbanowski *et al.*, 2000; Moller *et al.*, 2002b), although it is not necessarily essential, and several sRNAs show no interaction with this chaperone protein (Livny, 2012). The common functional elements of *trans*-acting sRNA are shown in Figure 1.9.

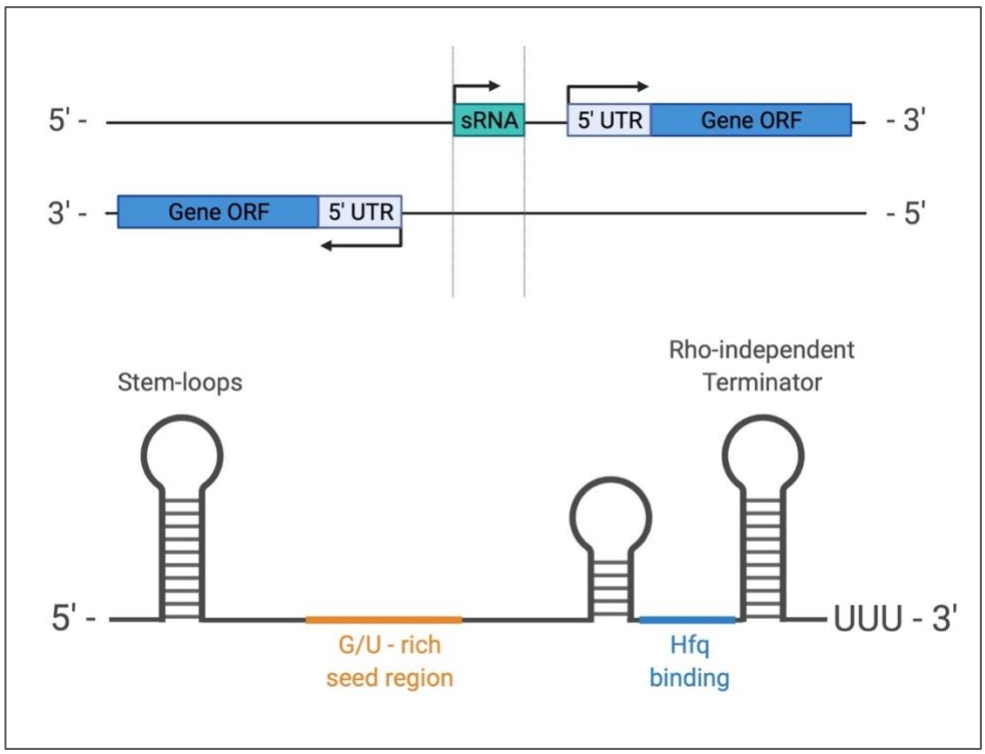


Figure 1.9. Common functional elements in *trans*-acting sRNAs. Diagram of a *trans*-encoded sRNA secondary elements consisting of stem-loops, the G/U rich interaction/seed region (yellow), a possible A/U-rich Hfq binding site (blue), a Rho-independent terminator, and the Poly(U) region at the 3' end (Gottesman & Storz, 2011; Svensson & Sharma, 2016).

1.2.2. Molecular mechanisms of regulation by sRNAs

As mentioned above, the fundamental function of sRNAs is the regulation of protein expression, which is mediated by sRNA base pairing interactions with target mRNAs (Lease & Belfort, 2000; Vogel & Papenfort, 2006; Vogel & Luisi, 2011; Henderson *et al.*, 2013). sRNA regulation can link to targeted mRNA degradation, as well as non-degradation mechanisms (reviewed in Vogel & Luisi, 2011) as discussed below.

1.2.2.1. Non-degradation mechanisms

Initially, pairing of an sRNA with its target mRNA can physically obstruct the RBS from ribosome binding (Figure 1.10A), preventing translation initiation by the 30S ribosome subunit and, consequently, down regulating protein synthesis (Vogel & Luisi, 2011). Alternatively, sRNA-mRNA pairing can cause conformational changes in the mRNA secondary structure to indirectly block the RBS (Kaminishi *et al.,* 2007; reviewed in Vogel & Luisi, 2011). This occurs when the sRNA binds on a position distant from the RBS while inducing a change in structural conformation which indirectly affects access to the RBS (Figure 1.10B). A versatile example of such sRNA regulation is by the *trans*-acting sRNA GcvB in *Salmonella*. GcvB has been shown to inhibit translation of targets *oppA* and *dppA* by base-pairing on functional regions of these mRNAs, occluding the RBS to block ribosome access (Urbanowski *et al*. 2000). The same sRNA has shown to repress translation of the mRNA *gtlI* by binding on sequences distant from the RBS (Sharma *et al*., 2007). Regulation by GcvB is further discussed in section 1.2.4.

Although less common, another mechanism of gene regulation by sRNAs is the activation of mRNA translation by a change in its secondary structure conformation, to expose the RBS for recognition and binding by the ribosome to initiate transcript translation (Figure 1.10C). This mechanism can further be associated with enhanced mRNA stability, through occlusion of ribonucleases cleavage sites, thus avoiding degradation by ribonucleases (Lease & Belfort, 2000). A long-known, well characterised, example of translation activation by sRNA is the regulation of the *rpoS* mRNA, encoding the σ factor

38, expressed in *E. coli* under stress conditions. Upon interaction with the sRNA DsrA, a RNase cleavage site in the *rpoS* mRNA is obstructed, stabilising the mRNA transcript, and providing protection against ribonucleolytic attack, thus supporting translation (Lease & Belfort, 2000; Majdalani *et al*., 2002).

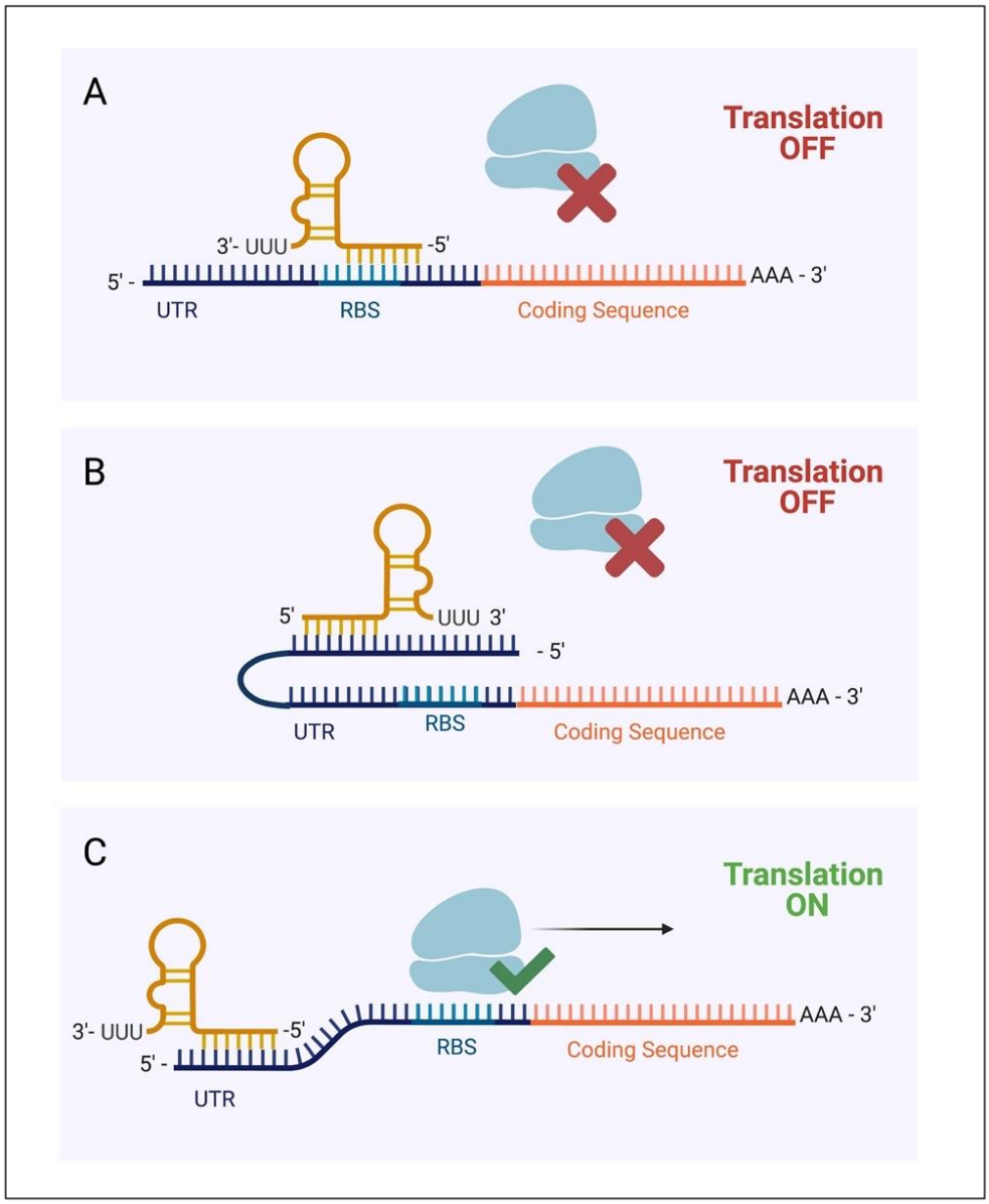


Figure 1.10. Mechanisms of sRNA regulation of mRNA translation. Targeting of mRNAs (5'UTR - dark blue, RBS - light blue and ORF - orange) by sRNAs (yellow) may lead to activation or inhibition of translation. **A)** sRNA-mRNA interaction (with or without Hfq) may lead to translational repression as the sRNA (yellow) may bind near or on the ribosomal binding site (RBS) (light blue) in the mRNA 5' UTR (dark blue) and block access of the ribosome. **B)** sRNA-mRNA interaction may change the mRNA secondary structure conformation, blocking access to the RBS by the ribosome **C)** In contrast, conformational change in structure may expose the RBS, facilitating the access of the ribosome to the mRNA, increasing the efficiency of the translation process (adapted from Vogel & Luisi, 2011).

1.2.2.2. Degradation mechanisms

RNA degradation is an essential component of cell metabolism and plays an important role in regulating transcript levels (Deutscher, 2006). Interaction with sRNAs can affect the stability and half-life of mRNAs by recruiting ribonucleases (RNases) resulting in the complete degradation of both (Georg & Hess, 2011; Lawal *et al*. 2011, reviewed in Vogel & Luisi, 2011). An example of sRNA-mediated translation inhibition by degradation is the sRNA RyhB, expressed in response to iron depletion to regulate synthesis of the iron-dependent superoxide dismutase, encoded in the *sodB* mRNA.

When interacting with the sRNA RyhB, the cleavage site is exposed for degradation by RNAse E, resulting in down regulation of the iron-dependent *sodB*, while up-regulating other iron-independent dismutases (Masse *et al*., 2003). Further, a mechanism of sRNA-mediated degradation of target mRNA has been shown to involve the 5'monophosphate of the sRNA stimulating RNase E cleavage of paired mRNA (Bandyra *et al.*, 2012). Following endonucleolytic attack of the mRNA via RNase E, free 3' ends are exposed resulting in exoribonuclease (e.g., PNPase) degradation of the transcript (Deutscher, 2006; De Lay & Gottesman, 2011) (Figure 1.11) and, consequently, silencing of the transcript.

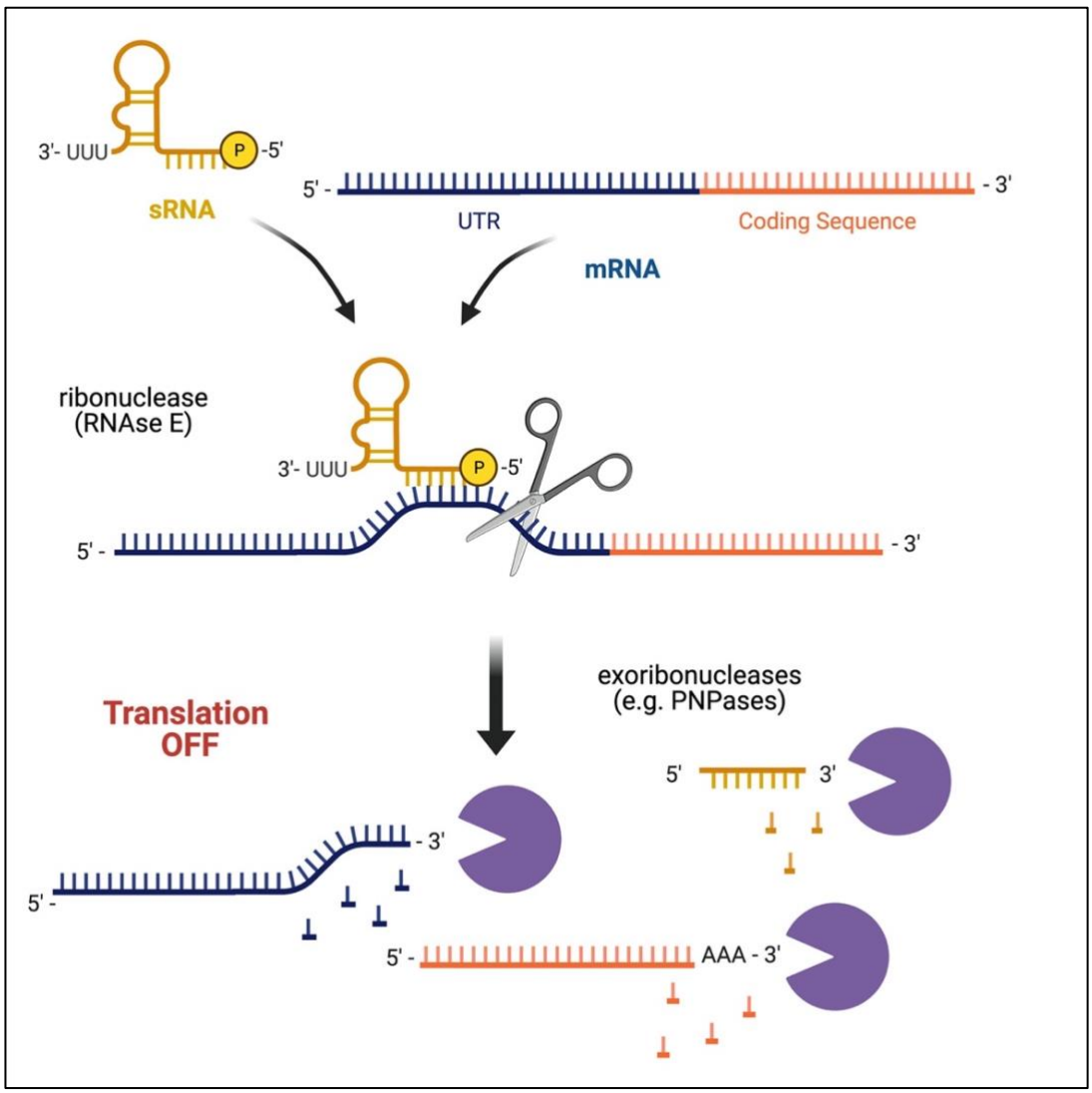


Figure 1.11. Schematic of an sRNA-induced mRNA degradation mechanism. Targeting of mRNAs (blue) by sRNAs (yellow) may lead to irreversible silencing through initial cleavage of the target molecule by ribonuclease E and then subsequent transcript degradation the via exoribonucleases such as PNPase (adapted from Bandyra *et al.*, 2012).

1.2.3. Role of sRNAs in virulence-associated metabolic pathways

As bacteria need to tightly regulate the expression of their virulence genes to adapt to the microenvironments where they are forced to survive, post-transcriptional gene regulation (PTGR) by sRNAs has been demonstrated to play a central role in orchestrating the synthesis of proteins associated to bacterial virulence by regulating their mRNA targets via base pairing mechanisms previously discussed in this chapter (Svensson & Sharma, 2016).

1.2.3.1. Virulence factors in bacteria

Bacteria are constantly exposed to environmental stimuli, such as variation in the availability of nutrients or the presence of toxic agents that cause stress in their cells. More specifically, pathogenic bacteria require an immediate metabolic response to changes in their host, as such impromptu adjustments are essential for their survival during colonisation and establishment of the infection. Generally, virulence factors are any molecules or proteins in bacteria that harm the host. Thus, virulence-associated metabolic pathways consist of molecular mechanisms producing virulence factors for pathogens to invade, colonize, survive, multiply and, consequently, cause disease in the host. Thus, bacterial virulence may be caused by over or under-production of certain proteins, and suppression of any of them can result in loss or reduction in virulence (Chiers *et al*., 2010; Svensson & Sharma, 2016; Sharma *et al*., 2017).

Virulence factors vary across species and most are yet to be characterised, however, these are associated with well-established metabolic pathways that pathogenic bacteria utilise to cause infection in the host, including: **adhesion to the host cells** during pathogenesis by producing factors such as fimbriae, a protein structure on the bacterial cell surface which binds to the host specific receptors to facilitate colonisation of the host, usually present in Gram-negative bacteria (Proft & Baker, 2009; Chiers *et al*., 2010; Svensson & Sharma, 2016); **acquisition of essential nutrients** by producing siderophores, which are receptors involved in iron acquisition (Diarra *et al*., 1996; Bossé *et al*., 2002; Chiers *et al*., 2010), and by synthesis of enzymes associated to amino acids biosynthesis pathways to adjust to nutrient deprived environments (Lone *et al*., 2009); **evasion of the host immunity** by secreting factors and proteases that break down the host cells by inducing phagocytosis to increase nutrients availability, while offering protection against the host macrophages, thus playing a role in infection persistence (Chiers *et al*., 2010; Svensson & Sharma, 2016), including the production of capsular polysaccharides (CPS), lipopolysaccharides (LPS) and enzymes involved in the processing of urea (Ward and Inzana, 1994; Bossé & MacInnes, 2000; Chiers *et al.*, 2010; Konieczna *et al*., 2012); and finally, **induction of lesions** by the release of exotoxins, with varied function across bacterial species, usually inducing cell death, with some possessing haemolytic and

cytotoxic properties, possibly damaging the host deeper tissues and cells to released invaluable nutrients and disabling of the host immune system (Frey, 1995; Bossé *et al*., 2014; Rudkin *et al*., 2017).

1.2.3.2. sRNAs linked to bacterial virulence

As noted above, sRNAs are linked to virulence in some bacteria. Indeed, sRNAs already mentioned in this chapter are linked to metabolic pathways associated with virulence and expressed in response to environmental changes. For example, the *E. coli* sRNA RyhB is expressed in iron depleted environments and shown to down regulate the synthesis of an iron-dependent enzyme encoded by *sodB* by degradation mechanisms involving ribonucleases (Massé *et al.*, 2003). Further, in *Shigella dysenteriae* RyhB was shown to be important for inhibiting *virB*, a transcription factor that encodes several virulence genes (reviewed in Matos *et al*., 2017). Additionally, in *Salmonella* cultivated in nutrient-rich medium, the sRNA GcvB is highly expressed and linked to inhibition of mRNAs of genes associated with amino acid synthesis and transport (Urbanowski *et al*., 2000; Sharma *et al*., 2007), suggesting involvement of this sRNA in the regulation of nutrient acquisition. Furthermore, in *Actinobacillus pleuropneumoniae* (APP), experiments by collaborators at Imperial College and University of Viçosa have shown that deletion of GcvB results in attenuation of virulence (Sanches, 2018). GcvB is conserved among widely characterised model organisms such as *E. coli* and *Salmonella* and, as the focus of this thesis, will be considered further below.

1.2.4. The sRNA GcvB

Amongst the widely characterised sRNAs, GcvB is one of the most well-known and conserved (Pulvermacher *et al*., 2008), and studies in model organisms *E. coli* and *Salmonella* have shown that GcvB has one of the largest regulons known in Gram-negative (Urbanowski *et al*., 2000; Sharma *et al*., 2007). In *Salmonella*, GcvB possesses a defined secondary structure containing several stem loops, a G/U-rich seed region and an Hfq-

binding site (schematics in Figure 1.12A), which might offer protection from immediate degradation (Sharma *et al*., 2007). GcvB plays an essential role in the regulatory system of genes that encode virulence factors associated with bacterial amino acid synthesis, including *IlvC*, *IlvE*, *ThrL* and *SerA* proteins, as well as the amino acid transport proteins, *DppA* and *OppA*, from *E. coli* and *Salmonella* (illustrated in Figure 1.12B) (Urbanowski *et al*., 2000; Sharma *et al*., 2007; Sharma *et al*., 2011; Miyakoshi *et al*., 2015).

Amino acid synthesis and transport is an essential metabolic pathway for establishing infection in the host by pathogenic bacteria. It provides substrates for the synthesis of proteins required in other metabolic pathways such as iron acquisition and biofilm formation, for example. These are highly expensive processes, which require efficient gene regulation, especially during bacterial virulence phases (Vitreschak *et al*., 2004). Unsurprisingly, GcvB is found in high prevalence within bacteria growing in nutrient-rich environments, and transcription is tightly regulated by the transcription factors GcvA and GcvR, as observed in both for *E. coli* (Urbanowski *et al*., 2000) and *Salmonella* (reviewed in Vogel, 2009). For these reasons, coupled with the observations of collaborators regarding GcvB attenuating virulence in APP, GcvB is an appealing candidate for therapeutic targeting, envisioning attenuation of virulence-associated metabolic pathways.

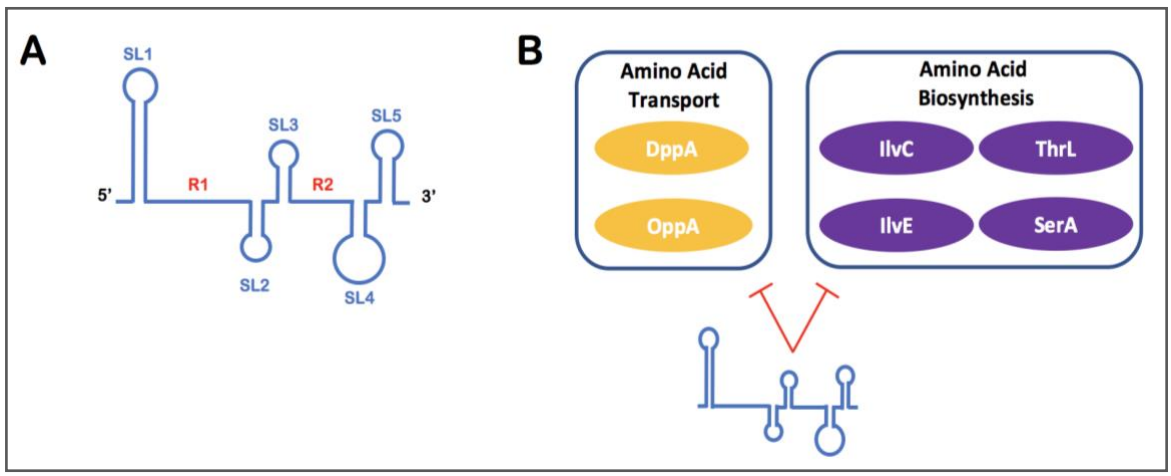


Figure 1.12. Schematic of GcvB structure and regulation. A) The *Salmonella* sRNA GcvB secondary structure contains five stem loops labelled as SL1-5, with SL5 predicted to act as a Rho-independent transcriptional terminator. The two consensus interaction regions are labelled in red as R1 and R2, with R1 base-pairing with target mRNAs while R2 interacts with the chaperone protein Hfq. These regions are highly conserved in other pathogenic bacterial species such as *E. coli* and *Vibrio cholerae* (Sharma *et al.*, 2007). **B)** GcvB sRNA represses the mRNAs of proteins involved in amino acid uptake (yellow) and synthesis (purple). The schematic information is based on studies of *Salmonella* GcvB (Sharma *et al.*, 2007) and *E. coli* GcvB regulation studies (Stauffer & Stauffer, 2005; adapted from Miyakoshi *et al.*, 2015).

1.2.5. Anti-sRNAs in bacteria

As non-coding RNAs have key roles in modulating the expression of target mRNAs, the question of how these elements regulated themselves naturally arises. In 2014, Tree *et al*. observed a novel class of non-coding RNAs in enterohemorrhagic *E. coli* termed anti-sRNAs which act by specifically binding to complementary regions in sRNAs resulting in their degradation. Interestingly, this study also revealed that the bacteriophage derived anti-sRNA AgvB binds to the R1 seed region in GcvB by mimicking an mRNA target, which antagonises the function of this sRNA (Tree *et al*., 2014). This mechanism is illustrated in Figure 1.13.

More recently, studies found the 3' UTR derived sRNA SroC base pairs with two exposed loop regions in GcvB, and that this interaction leads to RNase E-mediated degradation of the sRNA-sRNA pair (Miyakoshi *et al.*, 2015). It appears that nucleic acid binding/seed region recognition is an effective way in which bacteria regulate the activity of sRNAs. Therefore, a sensible therapeutic strategy would be to target virulence-associated sRNAs in this manner; by using nucleic acid mimics such as peptide nucleic acid (PNA) designed to disrupt the sRNA function. This approach is considered further in section 1.3.3.

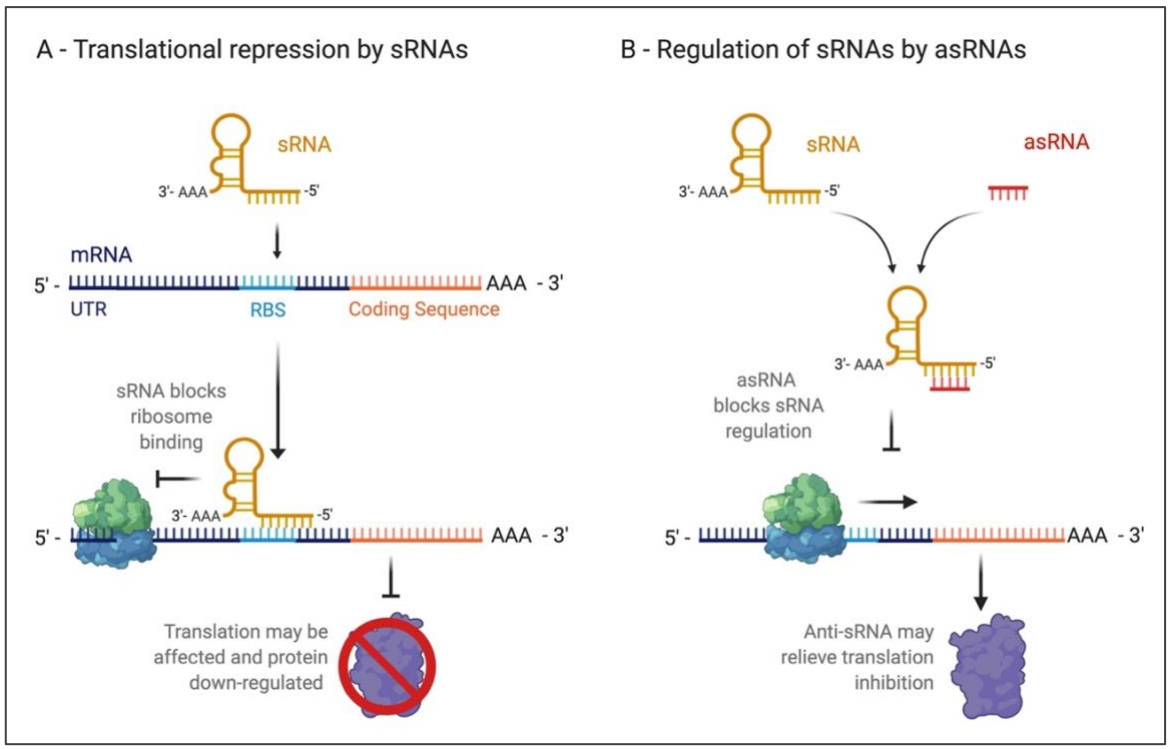


Figure 1.13. Schematic regulation of sRNAs by anti-sRNAs. A) Regulation of a mRNA target by a sRNA by base-pairing mechanism, for example, may repress protein synthesis. **B)** Bacteriophage-derived anti-sRNA antagonises the mechanism by binding to the seed region of the sRNA, liberating the mRNA for protein synthesis.

1.3. Porcine pleuropneumoniae caused by *Actinobacillus pleuropneumoniae* (APP) affects swine farming and the economy

Pig meat is considered the most consumed form of protein in the world, representing around 38% of the total consumption (U. S. Food and Drug Administration, 2016). In the United Kingdom, the average pork consumption is 26 kg per capita per year and forecasts suggests that pork consumption is on the rise (Agriculture and Horticulture Development Board UK, 2019). The growth in swine farming to meet market demand is correlated with high herd stocking densities, which promotes the propagation of infectious diseases, especially those of the respiratory tract. This is caused by animals being hoarded in restricted spaces to reduce costs and overall management. Sharing limited airspace in housing and transportation, combined with temperature changes, increase the opportunities for pathogen transmission (Chiers *et al*., 2010; Cho & Kim, 2011; Menzel *et al*., 2014). Endemic respiratory diseases in pigs result in high morbidity and mortality rates, elevating intervention costs and livestock losses. As a result, farming economy is negatively affected worldwide (Cho & Kim, 2011).

A 2018 review from WanderWaal and Deen on swine infectious diseases found that a considerable proportion of respiratory tract pathogens in pigs belongs to the *Pasteurellaceae* family. One such pathogen is *Actinobacillus pleuropneumoniae* (APP) (Shope, 1964), the etiological agent of the lethal swine disease porcine pleuropneumoniae, a highly contagious infection which mainly affects the lower respiratory tract. The disease can develop in a super-acute, acute and chronic form, depending on the host's immunity, the number of bacterial cells that reach the respiratory tract and the virulence of the causative agent of the infection. Once APP reaches the lungs, colonisation and production of toxins leads to severe tissue damage and possibly death of the animal within 48 hours (Hensel *et al*., 1995; Bossé *et al*., 2002). Currently, porcine pleuropneumoniae treatment includes the use of antibiotics such as tetracycline, sulphonamides, and penicillin. However, although antibiotics might control acute infections and transmission, these are shown to be ineffective in chronic occurrences and multi-bacterial infections (Vanni *et al*., 2012). Besides, constant usage of antibiotics contributes to the development of resistance and surge in new strains.

Furthermore, current vaccines are specific to each APP strain and do not prevent pathogenic colonisation (Marsteller & Fenwick, 1999; Bossé *et al*., 2002; Ramjeet *et al.*, 2008). In anticipation of an antibacterial resistance era, there is increasing interest in the development of novel therapeutic strategies to combat this disease. Therefore, studies involving the identification and therapeutic targeting of novel genes producing nucleic acids and proteins which are relevant to the physiology, regulation and virulence of pathogens can represent a starting point for the discovery of new alternatives for the control of swine pleuropneumonia.

1.3.1. Characteristics of APP

APP is a facultative anaerobic Gram-negative bacterium that belongs to the *Pasteurellaceae* family, which contains the genera *Haemophilus*, *Actinobacillus* and *Pasteurella*. The genus *Actinobacillus* consists of non-motile, non-spore-forming coccobacillus rods. This species presents beta haemolysis activity and is currently divided into two biological variants which differ in their dependence on nicotinamide adenine dinucleotide (NAD) for growth, with biovar 1 being NAD-dependent and biovar 2 NAD-independent. APP isolates are further differentiated in 18 serotypes, or serologically distinguishable strains, based on the polysaccharide antigens expressed on the capsule of the bacteria, with serotypes 13 and 14 belonging to biovar 2, and the others to biovar 1 (Bossé *et al*., 2002; Sthitmatee *et al*., 2003; Sárközi *et al*., 2015; Bossé *et al*., 2017; 2020).

In the onset of porcine pleuropneumoniae, APP cells escape immune response in the upper respiratory system of the host and reach the lower respiratory organs where they lysis a range of cells, including epithelial and red blood cells, causing lethal lesions in these tissues (Bossé *et al*., 2002; Rayamahji *et al*., 2005; Chien *et al*., 2009). Host cell lysis is crucial for APP to access nutrients from the environment and this species has developed intricate nutrient uptake mechanisms (reviewed in Chiers *et al*., 2010). Extensive lung damage is related to the stimulation of inflammatory cells by exotoxins and mainly to the combined effect of cytotoxins on different lung cells (Haesebrouck *et al*., 1997; Frey, 2003).

The injury of endothelial cells results in the activation of coagulation, formation of micro-thrombi and localised ischaemic necrosis (Bossé *et al*., 2002).

1.3.2. Virulence factors in APP

Virulence factors utilised by a specific pathogen can be conserved across different species or exclusive to its own. Although all APP serotypes are pathogenic, there are differences in virulence and disease severity. This discrepancy may be related to the diversity in expression of virulent factors among strains (Fuller *et al*., 2000; Lone *et al*., 2009; Klitgaard *et al*., 2010), including the quantity of endotoxins released into the host respiratory tissues such as capsular polysaccharides (CPS), lipopolysaccharides (LPS) and outer membrane proteins; production of adhesion factors used in biofilm formation; nutrient uptake capacity, and the production and excretion of pore forming exotoxins, which may cause lung lesions in the host (Lone *et al.*, 2009; reviewed in Chiers *et al*., 2010; Bossé *et al*., 2010; Xu *et al*., 2010).

1.3.2.1. APP exotoxins among serotypes

The major pathogenicity factor in APP is the production of the toxins Apx. These cytolysins are toxic to neutrophils and macrophages, causing cell death and intensifying the inflammatory response. The exotoxins ApxI, ApxII, ApxIII and ApxIV, members of the pore-forming RTX toxin group, are important factors involved in virulence. While ApxI is strongly haemolytic and cytotoxic, ApxII is weakly haemolytic and has moderate cytotoxicity. ApxIII is not haemolytic, but it is strongly cytotoxic (Frey, 1995; Haesebrouck *et al*., 2004; Chung *et al*., 2007; Opriessnig *et al*., 2011). All APP serotypes produce one or two of these toxins, which have been well characterised (see Table 1.1) (Chiers *et al*., 2010). In general, serotypes that produce ApxI are highly virulent, while serotypes producing ApxII and ApxIII present weak to moderate virulence. All strains produce the ApxIV protein *in vivo*, and its gene is used as an APP-specific biomarker in diagnostic tests (Gottschalk, 2003; Bossé *et*

al., 2014). APP serotype 8 is the most predominant in the UK and is considered of moderate virulence (O'Neill *et al*., 2010). Serotyping of APP is based on the CPS.

Table 1.1.
Example of exotoxins (Apx) and virulence profiling in fifteen different serotypes.

Strain	Biovar	ApxI	ApxII	ApxIII	ApxIV	Virulence
APP serovar 1 str. 4074	1	x	x		x	High
APP serovar 2 str. S1536	1		x	x	x	Moderate
APP serovar 3 str. S1421	1		x	x	x	Moderate
APP serovar 4 str. M62	1		x	x	x	Moderate
APP serovar 5b str. L20	1	x	x		x	High
APP serovar 6 str. Femo	1		x	x	x	Moderate
APP serovar 7 str. WF83	1		x		x	Weak
APP serovar 8 str. 405	1		x	x	x	Moderate
APP serovar 9 str. CVJ13261	1	x	x		x	High
APP serovar 10 str. D13039	1	x			x	Weak
APP serovar 11 str. 56153	1	x	x		x	High
APP serovar 12 str. 1096	1		x		x	Weak
APP serovar 13 str. N273	2		x		x	Weak
APP serovar 14 str. 3906	2	x			x	High
APP serovar 15 str. HS143	1		x	x	x	Moderate

The table contains extensively studied APP serotypes (1 to 15) and their Apx content. The strains considered of high virulence mainly express ApxI proteins. This occurs because the ApxI toxin is strongly haemolytic and cytotoxic. Strains of weak to moderate virulence express a combination of ApxII and ApxIII. ApxII is weakly haemolytic and present moderate cytotoxicity, while ApxIII is not haemolytic, but it is strongly cytotoxic.

1.3.2.2. Other mechanisms of virulence in APP

APP requires an adequate and fast reaction to changes, and while some aspects of its pathogenicity may be explained by the production of toxins, there are other expensive molecular systems, such as amino acid synthesis and transportation, that support the production of virulence factors as previously discussed. This ability of rapidly adapting to new conditions is orchestrated by a complex regulation in their molecular processes and pathways, including the manufacturing of a repertoire of proteins and enzymes, and it is possible that a part of these factors is regulated by sRNAs (Bossé, 2002; Sheehan *et al*., 2003; Lone *et al*., 2009; Rossi *et al*., 2016). Chiers *et al*. (2010) conducted a study to deepen the knowledge on APP virulence by collecting information on previously investigated genes with confirmed or putative involvement in virulence-associated pathways, and this information has been condensed and adapted in Table 1.2 below.

Table 1.2.
Virulence factors with confirmed or putative involvement in APP.

Virulence action	(Putative) Virulence factor	Gene
Adhesion	Type IV fimbriae (structural subunit 3 biogenesis components)	*apfABCD*
	Lipopolysaccharide biosynthesis	*galU, rmlC, rfbN,*
	Putative adhesin (OmpA outer membrane protein homologue)	*pomA*
	Fibronectin binding outer membrane protein	*comE1*
	Biofilm formation	*pgaA, pgaC, tadF, apfB*
	Putative fimbria-like protein (possibly involved in microcolony formation)	*flpD*
	Putative fibronectin binding	*tufA*
Induction of Lesion	Pore forming RTX toxin I, II, III and IV (activator, structural unit, secretion proteins)	*apxICABD, apxIICA, apxIIICABD, apxIVA*
	Lipopolysaccharide biosynthesis	*galU, rmlC, rfbN, rfbP, rfbU, rfaE*
Evasion Host Immunity	Cu-Zn superoxide dismutase	*sodC*
	Capsular polysaccharide	*cpxDCBA*
Persistence	Hemoglobin binding protein	*hgbA*
	Lipopolysaccharide biosynthesis	*galU, rmlC, rfbN, rfbP*
	Branched-chain amino acid biosynthesis	*IlvI*
Nutrient Uptake	Iron (chelated) ABC transporter, periplasmic binding protein	*yfeA, yfeB, yfeC, yfeD*
	Maltose regulon	*malEFG, malT, malPQ*
	Putative arginine/ornithine antiporter	*arcD*
	Arginine transport system permease protein	*artQ*
	Branched-chain amino acid transport system carrier protein	*brnQ*

(adapted from Chiers *et al*., 2010)

Although there are no published studies confirming regulation by sRNAs in APP, a set of novel sRNAs in APP have been identified during a previous computational study of APP serovar 5 strain L20 by Rossi and colleagues (2016), which are predicted to target the mRNAs of various genes in Table 1.2. Interestingly, few of these putative sRNAs match those already characterised in various widely known bacterial species, such as the sRNA GcvB (Vogel & Luisi, 2011). The direct targets of the putative GcvB in APP are unknown, as is the role of this sRNA in APP virulence, although unpublished data from Sanches (2018) suggests its deletion in APP results in attenuation of virulence. This information, coupled with the high sequence conservation in the key seed regions of the predicted GcvB, highlight that there is a possibility that the known functional properties of this sRNA in *E. coli* and *Salmonella* may translate to APP (Rossi *et al*., 2016). Indeed, whilst there is little known on sRNA regulation in APP, leaving a gap in the understanding of the regulatory networks for this species, the preliminary data for APP GcvB indicate its role in regulating virulence. APP GcvB could therefore represent a potential target for inhibiting virulence through antisense approaches for therapeutic benefit.

1.3.3. Exploiting therapeutic targeting of sRNAs via an antisense approach

With antibiotic resistance on the rise, there is a clear and growing need for new antibacterial strategies targeting alternative and/or unconventional mechanisms. The key role of sRNAs in pathogenesis and regulating virulence identifies them as potential targets for antimicrobial chemotherapeutics, both for infections of animals as well as humans (Dersch *et al*., 2017). Furthermore, with the rapid growth of RNA therapeutics and antisense technologies over recent years, targeting sRNAs via an antisense approach is now possible. Antisense approaches usually involve single-stranded nucleic acid mimic (NAM) molecules being designed to be complementary to an RNA target, such as an mRNA, in order to bring about a silencing effect through either translation blocking or facilitating degradation through creating a site for RNase cleavage (reviewed in Sully & Geller, 2016). In the case of targeting an sRNA, binding of a NAM would inhibit the sRNA's function as a means of bringing about the required outcome, such as inhibiting a virulence mechanism.

Advances in chemical modifications of NAMs such as phosphorodiamidate morpholino (MO), peptide nucleic acid (PNA), and locked nucleic acid (LNA) (Figure 1.14) have supported improvements in NAM molecule stability, as well as significantly enhancing its binding affinity for the RNA target (reviewed in Khvorova & Watts, 2017; Rasmussen *et al*., 2007). In addition, conjugation of the NAM to cell penetrating peptides has addressed challenges with cellular uptake (reviewed in Bai *et al.,* 2010). This has supported numerous studies demonstrating the potential of antisense approaches, using NAMs, towards targeting bacterial RNAs for antibacterial effect (reviewed in Wojciechowska *et al.,* 2020)

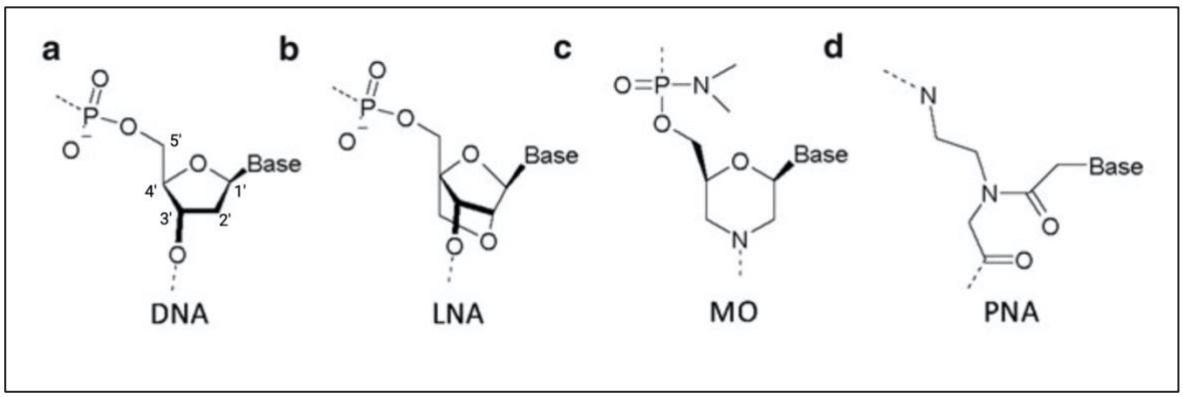


Figure 1.14. Chemical structures of NAM modifications. (a) DNA is shown as a comparison to modified NAMs used in antisense therapies. These include (b) locked nucleic acid (LNA), (c) phosphorodiamidate morpholino (MO) and (d) peptide nucleic acid (PNA) (adapted from Bertucci *et al*., 2012).

1.4. Identifying sRNAs and predicting their mRNA targets

The systematic study of complete genomes began with the use of DNA sequencing to expand genetic maps. In the case of prokaryotes, knowing the genes of a species can bring valuable information about its metabolic processes, including the genes involved in the production of virulence factors. However, RNA sequencing gives information on the DNA sequences that are actively expressed. The genes that encode proteins are identified by algorithms designed to recognise a set of signals in the DNA sequences that characterise an open reading frame (ORF), such as translation start and end codons or ribosome binding sites. However, these parameters are not properties of the genes encoding sRNAs, and due to their importance in prokaryotic gene regulation, the pursuit of novel sRNAs has become a rapidly growing research area, requiring novel approaches of analysis to support their identification within genomic and transcriptomic data.

Genome sequence assembly, both *de novo* and genome-guided, are essential for the study of sRNAs. The mechanisms by which sRNAs regulate gene expression are imprinted in the transcriptome of the cell. Next Generation Sequencing (NGS) is considered as a most powerful methodology to explore sRNA transcriptomes of bacteria growing in specific physiological conditions. This process of RNA sequencing (RNA-seq) is often used to catalogue all RNA transcripts, such as mRNA, rRNA, tRNA and sRNA (Hrdlickova *et al*., 2017). Indeed, transcriptomic studies of different organisms have revealed the existence of hitherto unknown transcripts and novel classes of sRNAs (Le Rhun *et al*., 2016).

Initially, computational methods to predict sRNAs applied substitution patterns to identify conserved DNA sequences present in intergenic regions (IGRs) of phylogenetically related species. These also searched for promoter sequences and binding site for conserved transcription factor and Rho-independent terminator in intergenic regions (Rivas *et al.*, 2001). Bioinformatic tools such as PePPER (de Jong, 2012) and TransTermHP (Kingsford, 2007), are used to identify clues alluding to transcriptional start and terminator sites located within a 50-600 nucleotide region. Over time, prediction approaches started to incorporate algorithms for gene annotation, base composition statistics, and secondary

structure prediction in non-conserved RNA sequences and thermodynamic stability (Pichon & Felden, 2008).

Colleagues at Imperial College London and University of Viçosa Brazil published a list of potential sRNAs in APP serovar 5 L20 (Rossi *et al*., 2016), generated by a computational approach using a combination of four prediction algorithms (RNAz, SIPHT, INFERNAL and searches in Rfam). An experimental approach using RT-PCR, Northern Blotting and RNA-seq data was carried out to validate the putative sRNA sequences contained in the list. The putative sRNA sequences were then used to predict any mRNA partners that they may interact, generating a map with possible sRNA-mRNA interactions. Although these methods validate the presence of sRNA and mRNAs in APP, the interaction between these pairs are yet to be validated.

1.5. A novel approach to testing putative sRNA-mRNA interactions

Understanding the mechanistic details of the regulatory activities of RNA is critical for understanding both normal cellular function and disease. Given the potential numbers of interactions involved, this in turn requires high-throughput techniques that are capable of detecting RNA-based interactions and analysing their functional outputs at the molecular level. Array technologies, in which can allow thousands of miniature assays can be performed in parallel on a single surface, has immense potential in this regard. RNA arrays are in their infancy compared to DNA arrays but, recently, a number of methods have been developed for generating RNA arrays (Lietard & Somoza, 2019).

The Callaghan group (University of Portsmouth) has developed a novel method for generating functional-RNA arrays through *in vitro* transcription from a DNA *in vitro* transcription (IVT) template array and *in situ* surface capture of the RNA on a facing surface (Phillips *et al*., 2018). As shown schematically in Figure 1.15, this method first involves producing a DNA *in vitro* transcription template array by spotting custom-designed DNA IVT templates onto a microarray slide. A DNA IVT template array – *in vitro* transcription reagent mix – RNA capture surface "sandwich" is then assembled. As *in vitro* transcription proceeds, RNA synthesised from each DNA IVT template is captured *in situ* on an RNA capture surface to generate a corresponding functional-RNA array. Applications such as detecting RNA-RNA (Phillips *et al*., 2018) and small molecule-RNA (Phillips *et al*., 2018, Henderson *et al*., 2019) interactions, and monitoring regulatory outputs such as protein expression (Norouzi *et al*., 2019) have been explored. Furthermore, the approach could be useful for situations where a particular RNA has many RNA-binding targets. This is recognised as the case for a number of *trans*-acting sRNAs, where they are known, or have been proposed, to interact with many mRNA targets. In this situation, a functional-RNA array of mRNA binding targets could be probed with a single sRNA, thus providing a high-throughput approach to rapidly testing sRNA-mRNA interactions.

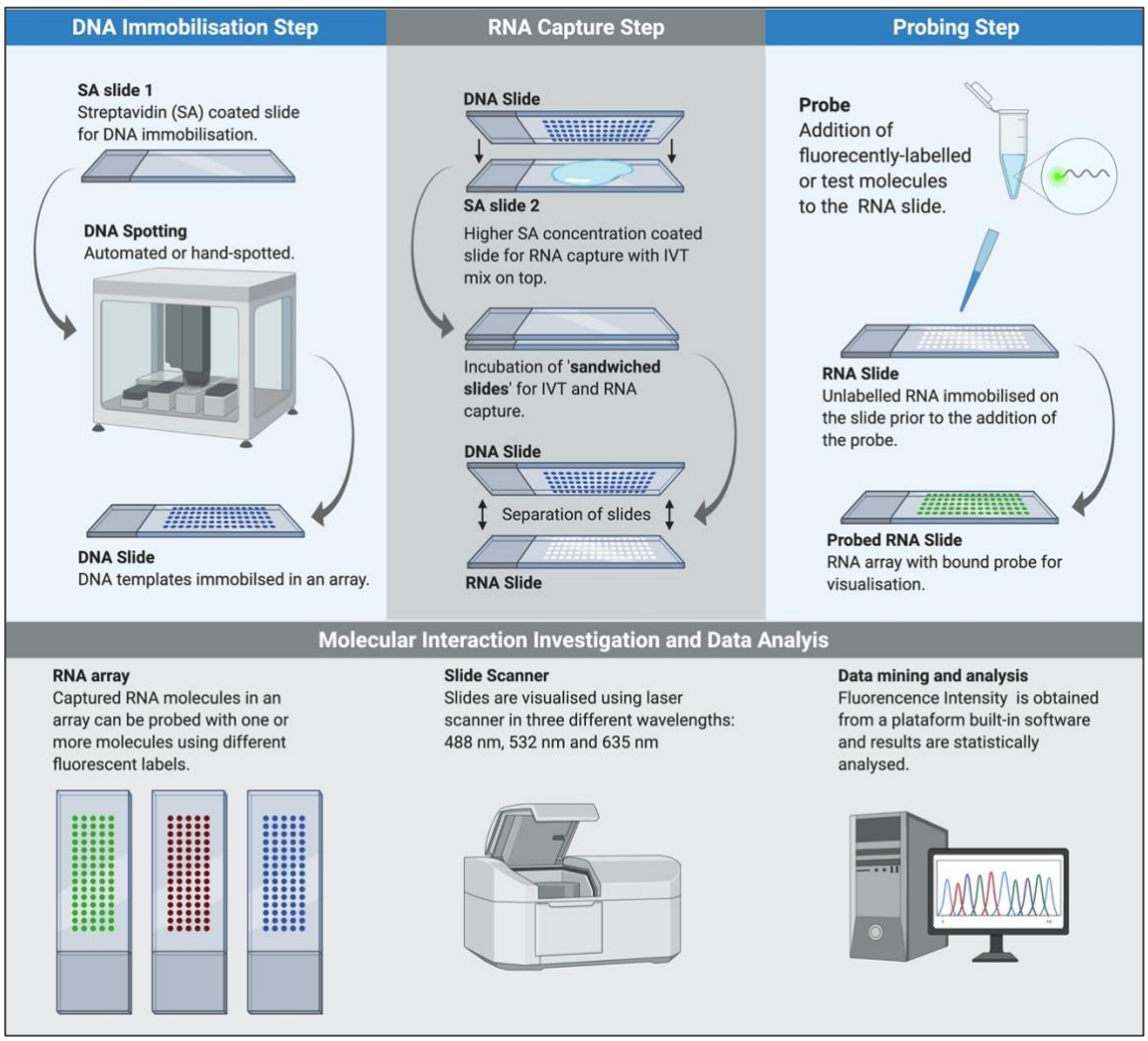


Figure 1.15. Schematics of the RNA array technology. The **DNA immobilisation step** involves biotin-tagged custom designed DNA *in vitro* transcription templates being spotted onto an SA-coated slide. The DNA slide is sandwiched together with a second SA-coated slide (coated with a higher concentration of SA), and with *in vitro* transcription (IVT) mixture in the middle. As the 'sandwich' is incubated, IVT proceeds, and the RNA is transcribed. As well as containing a specific RNA of interest, the transcribed RNA contains an SA aptamer (Philips *et al*., 2018) which support the **RNA capture step** in which the RNA is captured on the second SA coated slide, creating an RNA array slide. Following separation from the template DNA slide the RNA slide can be probed. The **probing step** involves the addition of fluorescently labelled control or test molecules which can then be visualised. Overall, the approach supports **molecular interaction investigation and data analysis**, specifically RNA arrays can be probed with differently labelled molecules, assessed in s slide scanner to give data for mining and analysis to understanding molecular interactions with RNA.

1.6. Project Aims

This research sought to explore sRNA-mRNA interactions linked to virulence in APP. With a specific focus on GcvB, the aims included:

1 – *In silico* studies to create a robust and adaptable bioinformatic pipeline to identify and analyse the expression of novel trans-acting sRNAs and their targets in APP serovar 8 MIDG2331 which are linked to virulence, and to select one novel sRNA candidate (namely GcvB) and its putative mRNA partners to be investigated using the RNA array technology.

2 – *In vitro* studies to validated and analyse GcvB-associate interactions using the novel RNA array technology.

3 – *In vitro* studies to exploit a strategy of interaction disruption between GcvB and its confirmed mRNA partners by designing and testing a nucleic acid mimic complementary to the GcvB seed region.

Chapter 2

2. Materials and Methods

2.1. *In silico* studies: Identification of novel trans-acting sRNAs involved in gene regulation in APP

Bacterial small non-coding RNAs have been widely characterised in model organisms such as *E. coli* and *Salmonella enterica* (Vogel & Luisi, 2011). However, these versatile functional molecules are yet to be characterised in *Actinobacillus pleuropneumoniae* (APP). In Chapter 3 of this study, a combination of bioinformatics tools was utilised to predict and analyse sRNA-mRNA pairs. The analysis included RNA-seq from APP serovar 8 MIDG2331 provided by collaborators Dr Janine Bossé and Professor Paul Langford from Imperial College London, generated as described in the subsections below.

2.1.1. Using transcriptomes to identify novel sRNAs

RNA sequencing (RNA-seq), commissioned by collaborators at Imperial College London, was used to identify novel sRNA candidates, and to validate the presence and expression levels of their putative target mRNAs, in APP serovar 8 MIDG2331 grown in different conditions. The set of data used in this study originated from four separate RNA-seq experiments. In the first strategy, the Ion Torrent platform was used to generate RNA-Seq reads for samples from wild-type MIDG2331 and an isogenic Hfq::3xFLAG mutant, grown aerobically and anaerobically in Brain Heart Infusion (BHI) broth, prepared with and without co-immunoprecipitation (Co-IP) using an anti-FLAG antibody to identify Hfq-binding sRNAs. For the other three experiments, referred to as Vertis 1-3, RNA-sequencing was performed by Vertis Biotechnologie AG (using the Illumina NextSeq 500 platform) for analysis of RNA samples prepared from MIDG2331 grown under different conditions. For

Vertis 1, whole transcriptome data was generated as strand specific 1x75 bp reads of rRNA-depleted RNA samples prepared from MIDG2331 cultivated on BHI agar plates incubated under aerobic (+AER) and anaerobic (-ANA) conditions. For Vertis 2, similar whole transcriptome data was generated for rRNA-depleted RNA samples prepared from MIDG2331 cultivated aerobically on BHI plate (WT), in BHI broth (BR), BHI broth with added 0.125 g/L ampicillin (AMP), and BHI broth plus 2g/L tylosin (TY), as well as from a delta-Hfq mutant of MIDG2331 cultivated in BHI broth (DHfq). Additionally, for each of these five RNA samples, a second library was prepared to allow specific enrichment of transcriptional start sites (TSS). For Vertis 3, whole transcriptome data was generated for rRNA-depleted RNA samples prepared from MIDG2331 cultivated aerobically in BHI broth (RNA-BHI) and porcine serum (RNA-PS).

2.1.2. RNA extraction and sample preparation

RNA extraction from the APP serovar 8 clinical isolate MIDG2331 was performed by colleagues at Imperial College, London. A DNase digest was performed and rRNA depleted from total RNA preparations using RiboZero rRNA Removal Kit for bacteria (Epicentre). Synthesis of cDNA was then performed from the rRNA depleted samples. For TSS analysis in the Vertis 2 set of samples, existing 5’ monophosphates in supplied RNA samples were identified through addition of a distinct linker sequence (CTGAAGCT), Tag1. RNA sequences were treated with tobacco acid pyrophosphatase (TAP), which facilitates the differentiation of TSS from processed 5’ ends (Argaman *et al*., 2001). By hydrolysing the terminal 5’-triphosphate group, found only in primary transcripts, to 5’ monophosphate groups, a 5’ sequence tag (TAATGCGC), Tag2, was ligated to the 5’ monophosphate. This acts to enhance the signal from transcription start sites (TSS). Two distinct tags were therefore present for existing 5’ monophosphorylated RNA (Tag1), and newly formed 5’ monophosphate RNA (Tag2). RNA was fragmented with ultrasound and cDNA synthesised, before ligation of 5’ and 3’ sequencing adaptors to antisense 1’strand cDNA. PCR amplification of cDNA was then performed and the TSS identified by the presence of Tag2.

2.1.3. Analysis of the RNA-seq data

RNA-seq files were uploaded to the NCBI-Short Read Archive (SRA) under the experiment SRX810211. The resulting bam files were analysed for quality control and processing of reads, mapped to the APP serovar 8 MIDG2331 genome, and set up for visualization in JBrowse (Skinner *et al*., 2009) in collaboration with the Hinton Group at the University of Liverpool. An average of 95% of the reads were successfully mapped to the genome. Full RNA-seq data analysis was carried out by Dr Will Rowe and Prof Jay Hinton at the University of Liverpool. Data and reports are available in the Hinton Group AWS server: https://s3-eu-west-1.amazonaws.com/hinton-analysis/projects/app-2017/APP-report.html

2.1.3.1. Software and approaches for analysing RNA-seq expression

Analysis of RNA-seq data was carried out using a combination of R packages and genomic software suites. The R statistical computing language and environment (R Core Team, 2017) provides a suite of packages for calculations on arrays, data analysis, including RNA-seq for either on-screen or as a file output. R Bioconductor (Gentleman *et al*., 2004) was used to analyse RNA-seq data in this thesis. The software suite BEDTools (Quinlan & Hall, 2010) was used to manually re-annotate the reference sequence genome LN908249.1 of APP serovar 8 MIDG2331 for purpose of this study only (Appendix 2). Statistical report on the RNA-seq data (see Appendix 5) was processed in the MultiQC software suite (Ewels *et al*., 2016).

2.1.3.2. Circos

Circos is a software package for visualising data and information in a circular layout. In this thesis, Circular Circos maps (Krzywinski *et al*., 2009) were generated to gauge the global representation of gene expression on the MIDG2331 chromosome and to compare expression in different bacterial growth conditions. The maps show coverage and relative

numbers of gene expression from RNA-seq data extracted from APP serovar 8 in different conditions, including BHI, and isolates in porcine serum medium (PS), added penicillin medium (AMP), added tylosin medium (TY) and anaerobic conditions (AN).

2.1.4. Comparative studies

The nucleotide sequence for the predicted novel small RNAs (sRNAs) of APP serovar 8 MIDG2331 used in this study was provided by collaborators at Imperial College, with their transcript confirmed by Northern blot and/or RNA-seq detection, using the computational approach developed by Rossi and colleagues (2016). All nucleotide sequences of sRNAs from other species and mRNA targets used in this study were extracted from the NCBI (National Center for Biotechnology Information) database (Sayers *et al*., 2019) and GenBank (Clark e*t al.*, 2016) using the accession number NZ_LN908249.1 for *Actinobacillus pleuropneumoniae* serovar 8 isolate MIDG2331.

The first part of the analysis was performed using the BLASTn algorithm (Basic Local Alignment Search Tool nucleotides), by querying the FASTA sequence of the sRNAs from APP serovar 8 MIDG2331 against the genome of selected species available in the NCBI database. The similarity (greatest positive%) and identity (greatest identity%) indices were used as criteria to infer about the conservation of the sequences.

The species used in this analysis were selected based on whether they were well characterised in literature, and their similarity pattern against ARRC01/RNA05. See Appendix 3 for a list of species and their accession information. Subsequently, the sequences for RNA05/ARRC01, within species strains and across phylogeny were aligned using Clustal Omega (Goujon *et al*., 2010), and visualised and labelled with the program JalView (Waterhouse *et al*., 2009).

2.1.5. Structural studies

The secondary structure of all the sequences used in this thesis were predicted using RNAfold (Gruber *et al.,* 2008), an online web tool for RNA modelling, prediction and analysis. The tool predicts minimum free energy (MFE) secondary structures and base pair probabilities from single sequences of RNA. Structures were predicted using option MFE and partition function option, RNA modelling parameters set for Turner model (2004), measured at 37^{o}C and remaining parameters kept as standard settings. Final image output choice was centroid structure drawing encoding base-pair probabilities, downloaded as PNG files and re-scaled accordingly.

2.1.6. Investigation of putative mRNA targets

2.1.6.1. TargetRNA2

The search for potential mRNA targets of the predicted sRNAs was performed using the program TargetRNA2 (Kery *et al.*, 2014), which identifies the putative sRNA binding positions in the target mRNA 5' untranslated region (UTR). Interactions between sRNAs and their putative mRNA targets scoring a p-value less or equal to 0.05 were selected, resulting in a list of target mRNA candidates for each sRNA input. The APP serovar 8 genome MIDG2331 used in this thesis was not available in the TargetRNA2 database at the time of this study. Instead, the algorithm uses the APP serovar 5 L20 genome as a reference for subsequent searches. The sRNA sequences from both serovars were checked for nucleotide mismatches in the homology studied in section 3.2.6, an only 1 mismatch was found at position 156 (Serovar 8 has A and serovar 5 has a G). This is outside GcvB interaction regions, thus unlikely to disrupt interaction with mRNA targets. Only annotated mRNA targets with identical nucleotide sequences and chromosomal position in both genomes have been considered in this study.

2.1.6.2. IntaRNA

Similar to RNATarget2, IntaRNA 2.0 predicts RNA-RNA interaction based on the sequence, structure and binding energy between the pair (Mann *et al*., 2017). However, IntaRNA allows for manual input of the mRNA sequences as Target (T) input and sRNA as query (Q), while RNATarget2 only allows for sRNA sequence input, producing the mRNA targets list automatically. At least two different species/strain sequences for the sRNA query are required. In this study the sequences for APP serovar 5 and serovar 8 were used. The number of interactions per RNA pair was set to 4 and the minimum number of base pairs interacting in the seed region was set as 7, the software standard.

2.1.6.3. STRING

STRING (Jensen *et al*., 2009) is an online database of both functional and physical relationships between proteins. Collating data held more widely from sources such as experimental data resources, computer driven prediction tools and public libraries, STRING acts as a central resource allowing the user to map all known interaction data together on to a common set of proteins and genomes. A list of genes from APP serovar 8 MIDG2331 predicted mRNA targets (Chapter 3, section 3.3.2) were analysed in STRING using the following parameters: a moderate confidence 0.400 and Markov clustering method (MCL) with inflation parameter 1.1.

2.2. *In vitro* studies: sRNA-mRNA validation and disruption assays using the RNA array

This section provides a list of materials and methods used in this study to investigate the putative sRNA-mRNA pairs predicted from *in silico* methods (Chapter 3) using the novel RNA array technology. The RNA array method was briefly introduced in Chapter 1, Section 1.5, and was used in Chapter 4 to validate the predicted sRNA-mRNA interactions as well as to exploit disruption of the validated interactions in Chapter 5. The standard protocol has been described previously (Phillips *et al*., 2018; Henderson *et al*., 2019; Vincent *et al*., 2021 *in press*) and details are provided below. Modifications in the protocol for assay optimisation are described in the corresponding results chapters.

2.2.1. RNA Array protocol

The novel RNA array technology has been briefly introduced in Chapter 1, Section 1.5. The concept involves an 'on array' *in vitro* transcription method for producing RNA arrays from template DNA arrays using a sandwich arrangement. However, the experimental considerations and practicalities required for generating functional-RNA arrays include a number of steps, shown schematically in the flow diagram in Figure 2.1 and detailed within the associated figure legend.

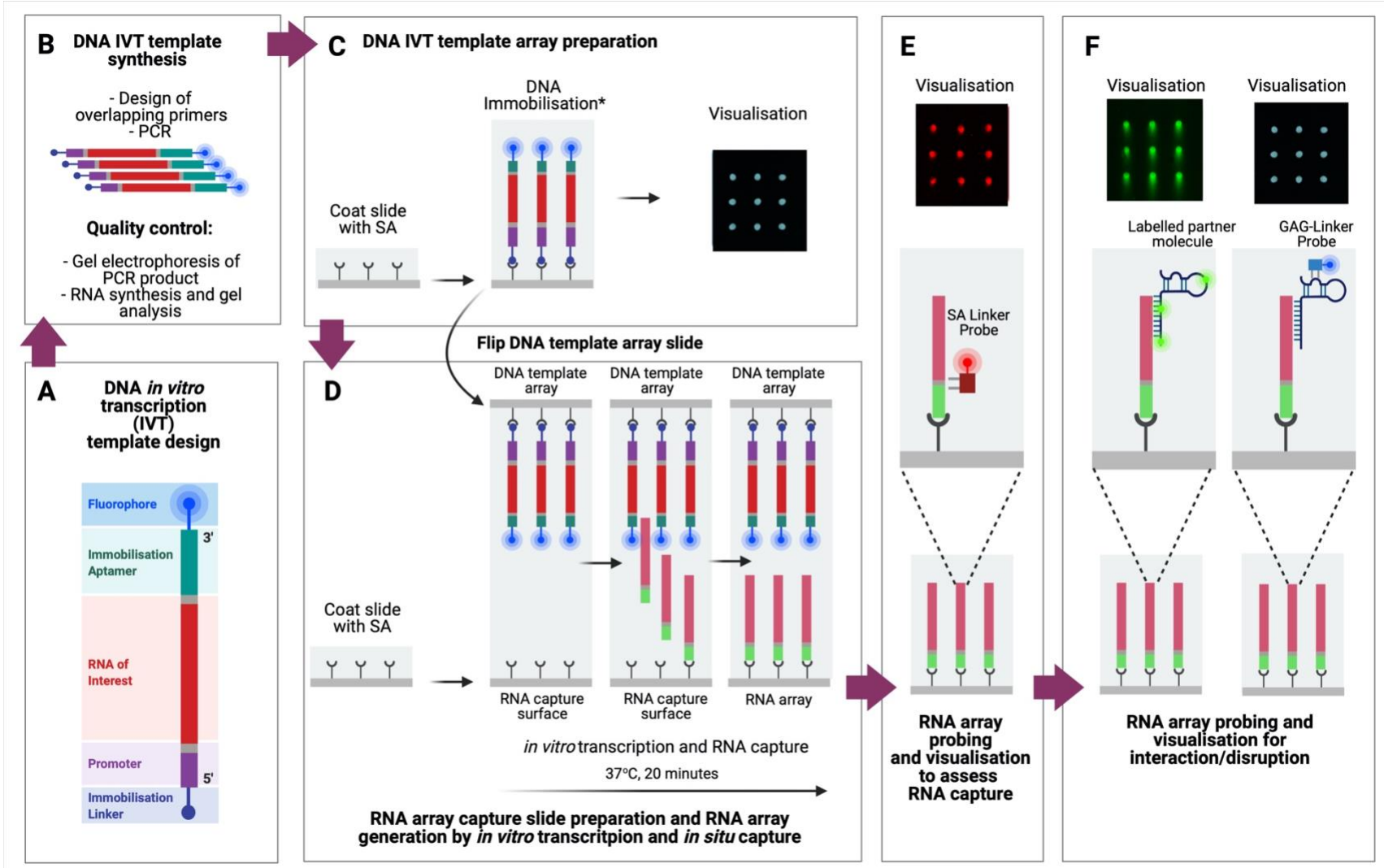


Figure 2.1. Flow diagram schematically illustrating the steps involved in generating functional-RNA arrays for use in interaction/disruption assays.

Figure 2.1. Continuation. A) Double-stranded DNA *in vitro* transcription templates are designed to include a T7 promoter (purple) with an optimum number of Gs to support transcription (grey adjacent to purple) and a sequence encoding an RNA of interest (mRNA regions in this study; red) coupled to a streptavidin RNA aptamer to support subsequent RNA immobilisation (green), separated by a flexible linker (grey), and tagged at their 5' end with a biotin-immobilisation linker and at their 3' end with an optional fluorophore. B) DNA *in vitro* transcription templates are synthesised following the overlapping extension PCR method (Gao *et al*., 2003). Quality control checks of the PCR products involves confirming correct size following gel electrophoresis, and in some cases, IVT of the PCR product followed by gel analysis to ensure the PCR product can be used as an effective template for IVT production. C) The DNA *in vitro* transcription template array is prepared by coating a Nexterion H (Schott) slide with 1µM streptavidin and then spotting the DNA *in vitro* transcription templates using either a pipette or automated arrayer robot. If an optional fluorophore has been included in the DNA template molecules, then the DNA template arrays generated can be visualised by scanning at the appropriate wavelength. D) RNA capture slide preparation involves coating Nexterion H (Schott) slide with 16.6 µM streptavidin and setting this up facing the DNA template slide in a 'sandwich' arrangement with *in vitro* transcription mix in the middle. As *in vitro* transcription proceeds, the RNA generated is immediately captured through the streptavidin aptamer, thus immobilising the RNA to the capture surface. Only full-length transcribed RNA is immobilised due to the inclusion of the streptavidin aptamer at the 3' end of the construct. E) The RNA array produced is probed with a DNA oligo, complementary to the streptavidin aptamer linker region, and labelled with an Alexa647 fluorophore (referred to as the SA-linker probe). The RNA array is then visualised using an excitation wavelength of 635 nm and a Standard Red emission filter to enable quantification of the RNA captured to be possible. F) For interaction/disruption assays, monitoring for efficiencies of partner molecule binding, the RNA array can be probed with either internally labelled binding partner RNA, or probed with unlabelled binding partner which is subsequently detected using a fluorescently labelled DNA oligo complementary to a specific region within the binding partner molecule (e.g., the Alexa488-labelled GAG-linker probe which binds to modified GcvB). Visualisation of the fluorophore enables assessment of partner molecule binding to targets on the RNA array.

*Please note that DNA templates can be spotted individually or as co-spotted mixtures of two *in vitro* transcription templates together. When transcribing from co-spotted DNA template arrays, for the pair of RNAs created one will include the SA-aptamer to support surface-immobilisation, whilst the second will only be linked to the RNA array if it interacts with the first RNA. More details of the co-spotting approach developed as part of this work is detailed in Chapter 4.

2.2.2. DNA *in vitro* transcription DNA template design

The architecture for design of the DNA IVT templates was followed as indicated in the standard RNA protocol (Phillips *et al*., 2018; Henderson *et al*., 2019; Vincent *et al*., 2021 *in press*). Generally, it includes: a biotinylated 5' linker for immobilisation on the DNA slide, the T7 promoter for on-slide IVT, the putative mRNA target DNA sequence for the RNA of interest (ROI), an additional linker (SA-Linker) adjacent to a sequence coding for a 3' streptavidin aptamer (SA Aptamer), which is suitable for post transcription on-surface immobilisation, and an optional 3' end linker with a specific fluorophore for detection and quantification (Figure 2.2). The sequences for DNA IVT template are in Appendix 9.

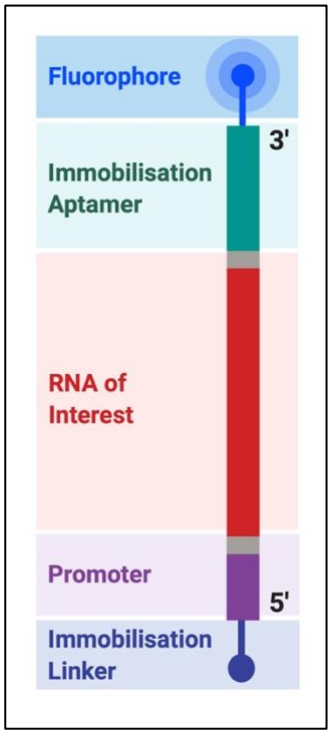


Figure 2.2. DNA *in vitro* transcription template design. A schematic of the DNA IVT template which consists of a short biotin-tagged immobilisation linker (dark blue), a promoter region (purple) with a transcription system-specific sequence (grey), sequence encoding the RNA of interest (red), an unstructured linker (light grey), sequence encoding an immobilisation aptamer (green) and an optional fluorophore (light blue).

2.2.3. DNA i*n vitro* transcription template synthesis

2.2.3.1. Design of overlapping oligonucleotides

Although all the DNA IVT templates were synthesised in-house by polymerase chain reaction (section 2.2.3.2), the overlapping oligonucleotides were ordered from Invitrogen (Thermo Fisher Scientific) and designed using a combination of software, including Microsoft Word and the Sequence Massager web-tool available from Biomodel website: http://biomodel.uah.es/en/lab/cybertory/analysis/massager.htm. Rules for the Thermodynamically Balanced Inside Out Polymerase Chain Reaction (TBIO-PCR) (Gao *et al*., 2003) methodology were observed, with oligonucleotide length set to a maximum of 90 oligonucleotides per overlapping oligo. This is slightly longer than the size used by Gao *et al.* (2003) to accommodate palindromic sequences contained in sRNAs and aptamers, as shorter oligos may increase unspecific binding of these sequences.

Flanking complementary regions in the 3' and 5' of overlapped each oligo, and the melting temperature for the overlapping regions varied from 58^{o}C to 64^{o}C using the standard formula [(C+G) x4] + [(A+T) x2]. Each overlapping oligo was provided in soluble powder and reinstated to 100 nM in nuclease free distilled water. DNA IVT templates required 4, 5 or 6 overlapping oligonucleotides. See Appendix 9 for oligonucleotides sequences.

2.2.3.2. Polymerase Chain Reaction (PCR)

Oligonucleotide overlap extension PCR incorporating the Thermodynamically Balanced Inside Out Polymerase Chain Reaction (TBIO-PCR) (Gao *et al*., 2003) methodology was used to construct and amplify DNA IVT templates. This technique is an improved version of the traditional Thermodynamically Balanced Conventional (TBC). It was selected over the simpler Overlap Extension (OE) methods due to the cost and time implications of using plasmid OE PCR. Other advantages of using *de novo* gene design and synthesis includes significant decrease in corrective mutagenesis, the possibility of easily adapt *cis*

elements in the gene of interest such as promoters, efficient addition of structural features to a functional RNA such as loops and aptamers, insertion of mutations or deletion of specific regions with no need for restriction enzymes cleavage sites.

The KOD Hot Start DNA Polymerase Kit (Merck) was used to carry out TBIO-PCR on the Applied Biosystems GeneAmp PCR system 9700 thermal cycler with Fisher brand 0.3 ml thin-walled PCR tubes. Assembly was carried out at room temperature with components thawed on ice. Component concentrations and thermal cycling parameters used for each DNA IVT template synthesis and amplification were set out in Table 2.1 for shorter sequences requiring 4 overlapping oligonucleotides (Cy3 labelled GcvB; Figure 2.3); Table 2.2 (Construct A; Figure 2.4) and Table 2.3 (Construct B; Figure 2.5) for longer sequences with 5 and 6 overlapping oligonucleotides respectively (mRNA targets); and Table 2.4 (Figure 2.6) for **M**odified **G**cvB (GcvB-**MG**) with 6 overlapping oligonucleotides.

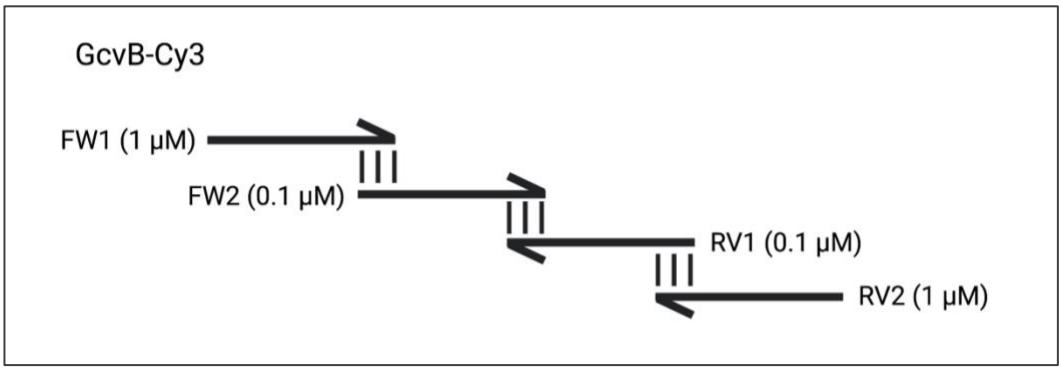


Figure 2.3. Illustration of the 4 overlapping primers used for TBIO-PCR for GcvB-Cy3.

Table 2.1.

Reaction mixture assembly for TBIO-PCR synthesis of the DNA template for GcvB-Cy3.

Item	Initial	Final	Volume (μL)
dH_2O	-	-	70
$MgSO_4$	25 mM	1.5 mM	6
KOD Hot Start Buffer	10 x	1 x	10
dNTPs	2 mM	0.2 mM	10
KOD Hot Start DNA Polymerase	1 U/μL	0.02 U/μL	2
FW1 Oligo	10 μM	1 μM	1
FW2 Oligo	1 μM	0.1μM	1
RV1 Oligo	1 μM	0.1μM	1
RV2 Oligo	10 μM	1 μM	1
Total			**100**

The reaction components were assembled in a microfuge tube and PCR thermal cycler parameters set as follow:

Temperature (°C)	Time	No. cycles
95	2’	1
95	10”	25
58	10”	
70	10”	
70	5’	1
4	∞	1

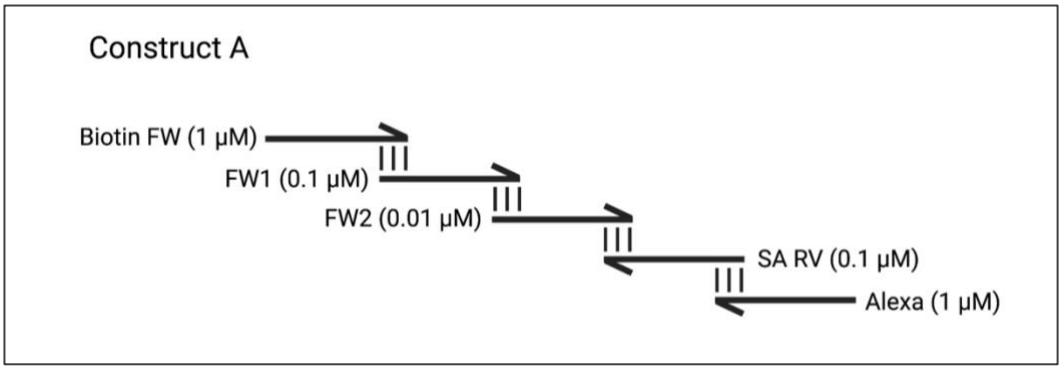


Figure 2.4. Illustration of the 5 overlapping primers used for TBIO-PCR for Construct A.

Table 2.2.

Reaction mixture assembly for TBIO-PCR synthesis of DNA template for mRNA targets using Construct A strategy.

Item	Initial	Final	Volume (µL)
dH_2O	-	-	67
$MgSO_4$	25 mM	1.5 mM	6
KOD Hot Start Buffer	10 x	1 x	10
dNTPs	2 mM	0.2 mM	10
KOD Hot Start DNA Polymerase	1 U/µL	0.02 U/µL	2
Biotin FW oligo	100 µM	1 µM	1
FW 1 Oligo	10 µM	0.1 µM	1
FW2 Oligo	1 µM	0.01 µM	1
SA RV Oligo	10 µM	0.1 µM	1
Alexa Oligo	100 µM	1 µM	1
Total			**100**

The reaction components were assembled in a microfuge tube and PCR thermal cycler parameters set as follow:

Temperature (°C)	Time	No. cycles
95	2’	1
95	10”	25
62	10”	
70	10”	
70	5’	1
4	∞	1

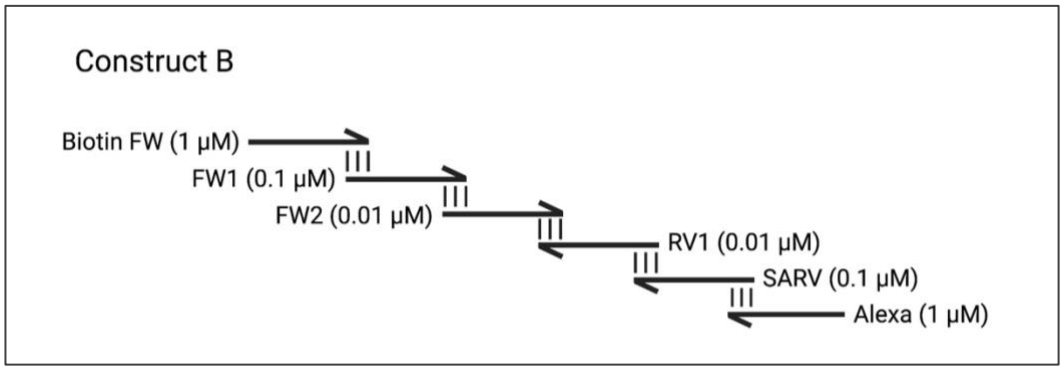


Figure 2.5. Illustration of the 6 overlapping primers used for TBIO-PCR for Construct B.

Table 2.3.

Reaction mixture assembly for TBIO-PCR synthesis of DNA template for mRNA targets using Construct B strategy.

Item	Initial	Final	Volume (µL)
dH_2O	-	-	66
$MgSO_4$	25 mM	1.5 mM	6
KOD Hot Start Buffer	10 x	1 x	10
dNTPs	2 mM	0.2 mM	10
KOD Hot Start DNA Polymerase	1 U/µL	0.02 U/µL	2
Biotin FW oligo	100 µM	1 µM	1
FW 1 Oligo	10 µM	0.1 µM	1
FW2 Oligo	1 µM	0.01 µM	1
RV1 Oligo	1 µM	0.01 µM	1
SA RV Oligo	10 µM	0.1 µM	1
Alexa Oligo	100 µM	1 µM	1
Total			**100**

The reaction components were assembled in a microfuge tube and PCR thermal cycler parameters set as follow:

Temperature (°C)	Time	No. cycles
95	2’	1
95	10”	35
60	10”	
70	10”	
70	5’	1
4	∞	1

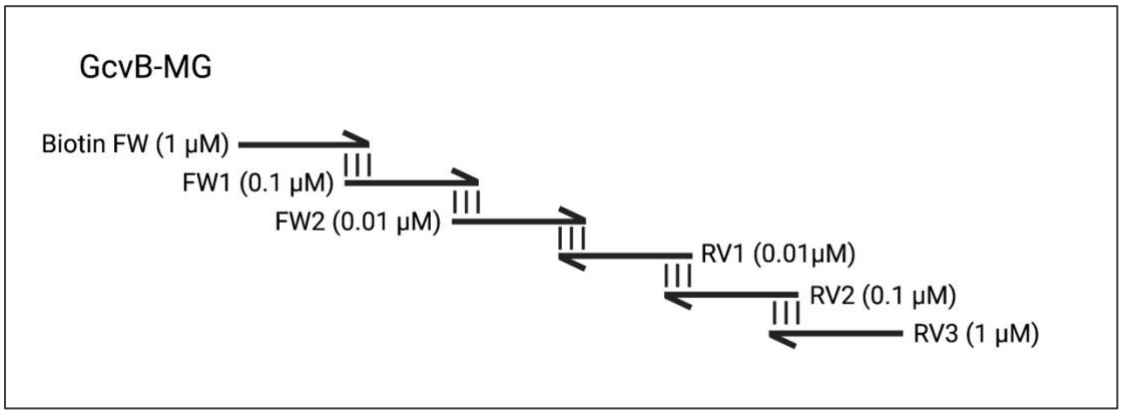


Figure 2.6. Illustration of the 6 overlapping primers used for TBIO-PCR for Modified GcvB (GcvB-MG).

Table 2.4.

Reaction mixture assembly for TBIO-PCR synthesis of the DNA template for GcvB-MG.

Item	Initial	Final	Volume (µL)
dH_2O	-	-	66
$MgSO_4$	25 mM	1.5 mM	6
KOD Hot Start Buffer	10 x	1 x	10
dNTPs	2 mM	0.2 mM	10
KOD Hot Start DNA Polymerase	1 U/µL	0.02 U/µL	2
Biotin FW oligo	100 µM	1 µM	1
FW1 Oligo	10 µM	0.1 µM	1
FW2 Oligo	1 µM	0.01 µM	1
RV1 Oligo	1 µM	0.01 µM	1
RV2 Oligo	10 µM	0.1 µM	1
RV3 Oligo	100 µM	1 µM	1
Total			**100**

The reaction components were assembled in a microfuge tube and PCR thermal cycler parameters set as follow:

Temperature (°C)	Time	No. cycles
95	2’	1
95	10”	35
58	10”	
70	10”	
70	5’	1
4	∞	1

2.2.3.3. PCR Purification

Purification of PCR products was carried using the Macherey-Nagel PCR clean up kit as described in the standard protocol. DNA was eluted by placing the column into a new tube, adding 15 μL of elution buffer, incubating at room temperature for 1 minute, and then centrifuged for 1 minute at 11,000 x g. The concentration of the purified PCR products was verified using the microvolume UV-VIS NanoDrop 2000 spectrophotometer, and their size and quality confirmed by gel electrophoresis run on a 1.8% (w/v) agarose gel (method described in section 2.2.3.4). See Appendices 11 and 12 for agarose gel of DNA IVT templates for GcvB, GcvB-MG and mRNA targets.

2.2.3.4. Gel electrophoresis analysis – Agarose gels

Gel electrophoresis is a method for the analysis of nucleic acids and proteins. It consists of a gel matrix where these biomolecules are separated on the basis of their size, shape, charge or all three. Depending on the purpose, the gel matrix and components can take many forms to provide a native or denaturing environment.

Agarose gels were used throughout as a quality control check to determine the size and quality of DNA products of PCR prior to RNA transcription. Due to the short size of the DNA templates, composition of each agarose gel was adjusted to a concentration of 1.8% (w/v), made with 1.8 g of agarose powder (Fisher) dissolved in 100 ml of 1 x TBE buffer (Appendix 14) by boiling the solution for approximately 2 minutes. Ethidium bromide was added to the dissolved solution in 0.5 μg/ml final concentration. The solution was poured into a casting tray with combs added to mould the wells. The gel was left set for 30 minutes before adding 500 ml of 1 x TBE running buffer and removing the combs. DNA samples were diluted to 100 ng in 8 μl of 1 x loading dye. Low Molecular Weight DNA Ladder was inserted to allow suitable calibration prior to running gels at 120 V for approximately 50 minutes. Gels were visualised under a Syngene UV transilluminator.

2.2.3.5. RNA synthesis by in vitro transcription (IVT) in a tube

Before being immobilized into the streptavidin-coated DNA IVT template slide, the DNA IVT templates of all mRNA target constructs were transcribed *in vitro* in a tube as a quality control check. To ensure transcriptional efficiency of designed array constructs, IVTs were carried out using the MegatScript T7 Transcription Kit (Thermo Fisher). A sample of each DNA IVT template was diluted accordingly to make up to 200 ng as instructed in their protocol. The IVT mixture was assembled in a microfuge tube for a 20 μL reaction:

Table 2.5.
IVT reaction mix for making unlabelled RNAs (MegaScript T7).

Component	Initial []	Final []	Volume (μl)
Nuclease-free dH_2O	-	-	To 20 μL
T7 Reaction Buffer	10 x	1x	2
DNA IVT Template	-	200 ng	-
rNTPs (each)	10 mM	1 mM	2
RNA Pol T7	10 x	1x	2

The microfuge tube was then incubated at 37^{o}C for 2 hours usually. Subsequently, the removal of DNA IVT templates was carried out by adding 2 μl Turbo DNAse to each tube and intubating for a further 10 minutes at 37^{0} C. Please note that a modified version of this method was used for preparing labelled RNA (section 2.2.7.1).

2.2.3.6. Gel analysis - Urea denaturing polyacrylamide gel electrophoresis (Urea-PAGE)

Following IVT, further denaturing of the secondary structures of RNA samples was achieved using urea polyacrylamide gel to verify whether the synthesised DNA IVT templates were transcribing the correct size transcripts. An 8% (w/v) denaturing Urea-PAGE gel was used to separate the nucleic acid bases of the RNA, allowing comparative measurement of molecular size. A gel mixture was prepared (Table 2.6) and poured in to a

Novex™ Mini gel cassette with combs inserted to produce sample wells. This was allowed to set for 30 minutes and the cassette was then clamped into the XCell SureLockTM Mini-Cell. A 1 x TBE running buffer was applied to the tank and after removing the comb the gel was pre-run at 20 Volts for 30 minutes to heat. The RNA samples were diluted to 100 ng in 6 µl of 1 x loading dye and heated to 95° C for 5 minutes prior to loading on to the prepared gel. The gel was run at 200 V for 90 minutes and removed from the cast for staining using SYBR Gold™ (Invitrogen). Gels were visualised under a Syngene UV transilluminator. For RNA fluorescently labelled with Cy-3 UTP, samples were diluted to 50 ng in 6 µl of 10% (v/v) glycerol without loading dye. Gels containing Cy3-labelled RNA were visualised under in a Fuji FLA5000 phosphorimager. See Appendix 13 for 8% (w/v) denaturing Urea-PAGE gel of IVT products for the DNA templates for the mRNA targets and GcvB-MG.

Table 2.6.
Denaturing urea-PAGE gel reagents (National Diagnostics).

Component	Volume
UreaGel™ concentrate	3.2 ml
UreaGel™ diluent	5.8 ml
UreaGel™ buffer	1 ml
TEMED	4 µl
10 % (w/v) APS	80 µl

2.2.4. DNA *in vitro* transcription template array preparation

2.2.4.1. Streptavidin coating of DNA template slide

Nexterion® Slide H (Schott Corporation) slides were stored in the -20°C freezer aliquoted inside a cartridge prior to experiment to avoid surfaces damaging. Slides were thawed at room temperature for 30 minutes prior to use. SA was diluted: 80 µl (1 µM) in 1 x PBS buffer (Appendix 14).

Slides were checked for dust and debris prior to SA coating and nitrogen gas was used to remove the debris. A volume of 80 µl of the diluted SA was pipetted in the middle of the slide, which was then covered with liftaslips. The slides were then incubated at 37^{o} C for 60 minutes in a humidified chamber, consisting of a Petri dish with 30 ml of Milli-Q dH_2O, to avoid drying the slide. After incubation was finalised, slides were submersed in amine blocking buffer (Appendix 14) for 5 minutes. The slides were washed to remove unbound streptavidin in the following consecutive steps:

- Submersed in 1 x PBST buffer for 5 minutes
- Extra wash in Milli-Q dH_2O for 30 seconds
- Final dip into Milli-Q dH_2O

The slide washing steps were performed in 45 ml volumes in Falcon tubes on a roller mill. The slides were then transferred to Falcon tubes to be dried in the centrifuge at 1000 rpm for 30 seconds. The slides were ready for the DNA immobilisation step.

2.2.4.2. DNA in vitro transcription template immobilisation on the slide

Immobilisation of DNA IVT templates on the slide was performed either manually by pipetting the sample onto the slide or by high throughput spotting using the QArray2 arrayer robot. The slide coated for DNA immobilisation (1 µM SA) was placed into the robot in a slide slot within a humidified chamber (55%). The samples containing the DNA IVT templated were transferred into wells in a 384-wells cassette. Each well containing a given DNA IVT template was annotated to be identified by the arrayer robot and later matched to the correspondent spot after fluorescence visualisation under the laser scanner. The typical concentration for the DNA spotting was between 5 nM to 450 nM per spot, and 130 spots required 1µl of DNA sample under optimal humidity with a minimum of 5 µl by well to avoid sample removal by evaporation. Once spotting was finished, the slides were immediately transferred to the 37^{o}C incubator and left for 20 minutes in a humidified

chamber. The slides were washed to remove unbound DNA IVT templates in the following the same protocol from the previous SA coating step:

- Submersed in 1 x PBST buffer for 5 minutes
- Extra wash in Milli-Q dH_2O for 30 seconds
- Final dip into Milli-Q dH_2O

After the washing steps the slides were dried in the centrifuge at 1000 rpm for 30 seconds and were ready to be used for generating RNA arrays following *in vitro* transcription (IVT).

2.2.4.3. Co-spotting two DNA in vitro transcription templates

The same method as described in 2.2.4.2 was carried out, except mixtures of two DNA IVT templates were prepared for spotting together. In these situations, it was usual for only one of the DNA templates to be fluorescently labelled.

2.2.4.4. Visualisation of the DNA in vitro transcription template array slide

DNA *in vitro* transcription template arrays were scanned using a GenePix 4300A microarray scanner with integrated GenePix Pro 7 image analysis software (Molecular Devices). The templates commonly incorporated three choices of dye: Dy-588, Dy-532 and Dy-647 fluorescent-label which was visualised with a microarray scanner using an excitation wavelength of 488 nm (Standard Blue filter), 532 nm (Standard Green filter) and 635 nm (Standard Red filter), respectively. The fluorescence intensity of each spot on the array was quantified using the integrated image analysis software GenePix Pro 7 on the microarray scanner.

2.2.5. RNA capture slide preparation and RNA array generation by IVT and *in situ* capture

2.2.5.1. Streptavidin coating of RNA capture slide

The same method as described in section 2.2.4.1 was carried out, except the SA concentration used was changed to 16.6 μM to create an RNA capture slide.

2.2.5.2. In vitro transcription and *in situ* RNA capture

In this step, the DNA IVT template array slide and RNA capture slide were 'sandwiched' together (Figure 2.1, D) with IVT mixture (Table 2.7) in the middle. The DNA IVT templates on the DNA template array slide were used to produce RNA which was then captured directly on the RNA capture slide. In terms of the set-up, the IVT mixture (25 μl) was pipetted onto the RNA capture slide (Figure 2.7). This avoids starting the IVT reaction prior to the 'sandwich' being formed; something which would occur had the IVT mixture been put on the DNA slide prior to 'sandwich' arrangement set up. The IVT mixture was spread throughout the entire RNA capture slide area and the 'sandwich' created. Once 'sandwiched', the slides were immediately transferred to the 37°C incubator for 20 minutes in a humidified chamber created with a Petri dish and 30 ml of Milli-Q dH_2O.

Table 2.7.
IVT reaction mixture for the IVT step in RNA array method (MegaScript T7).

Component	Initial []	Final []	Volume (μl)
Nuclease-free dH_2O	-	-	To 25 μL
T7 Reaction Buffer	10 x	1x	2
rNTPs	10 mM	1 mM	2
RNA Pol T7	10 x	1x	2

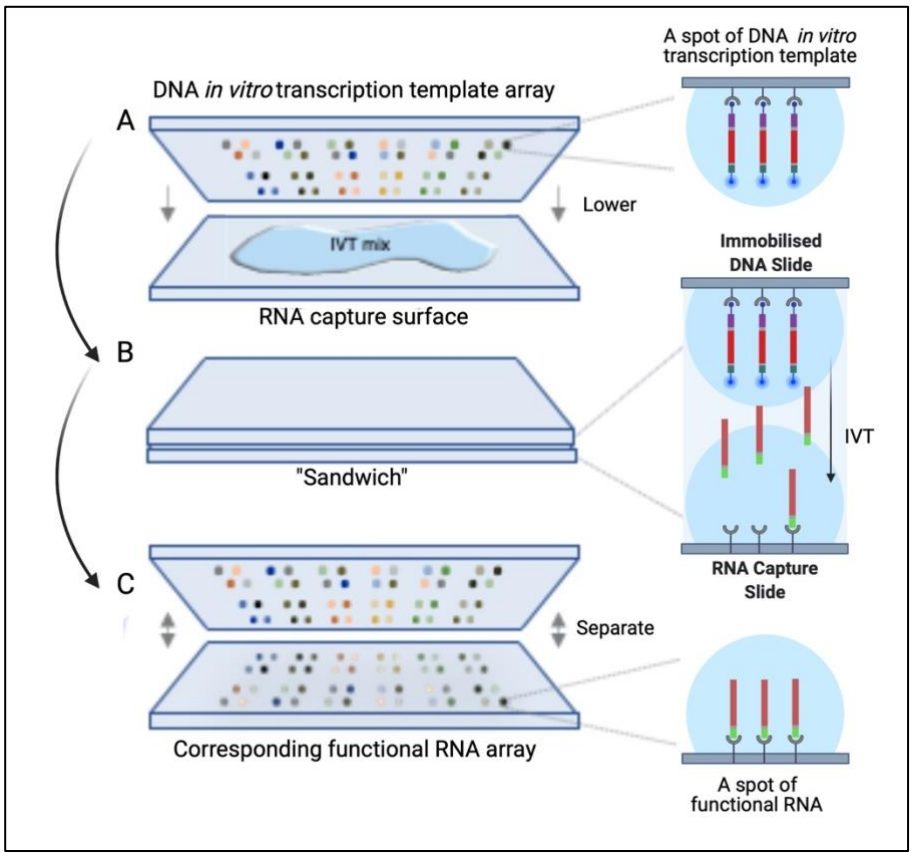


Figure 2.7. Illustration of the setup of the *in vitro* transcription and *in situ* RNA capture step. A) Assembly of the DNA *in vitro* transcription template array – *in vitro* transcription reagent mix (IVT mix) – RNA capture surface "sandwich". The RNA capture surface is position surface-side up. *In vitro* transcription reagent mix is pipetted onto the RNA capture surface and the DNA *in vitro* transcription template array, array-side down, is carefully lowered onto the RNA capture surface to complete the "sandwich" assembly. (B) The "sandwich" assembly is incubated at 37°C for 20 minutes during which time RNA is synthesised by *in vitro* transcription and captured *in situ* by the RNA capture surface. (C) The DNA *in vitro* transcription template array and the newly generated corresponding functional-RNA array are then carefully separated.

At the end of the IVT step, the slides were separated in Milli-Q dH_2O and immediately washed as follow:

- Submersed in 1 x PBST buffer for 5 minutes
- Extra wash in Milli-Q dH_2O for 30 seconds
- Final dip into Milli-Q dH_2O

After the washing steps the slides were dried in the centrifuge at 1000 rpm for 30 seconds and the transcribed unlabelled RNA immobilised in the RNA slide was ready to be

investigated via the RNA probing steps (section 2.2.6). Please note that for co-spotted DNA template array slides, the two RNAs of interest, considered as potential binding partners of each other, would be transcribed at the same time. The pair of RNAs typically included a target mRNA, appended with an SA aptamer which would surface-immobilise to the RNA capture slide, and a potential partner sRNA lacking the SA aptamer. Thus, the transcribed partner would only be identified on the RNA array via an interaction with its target RNA, which was surface-immobilised to the RNA capture slide. This method was developed as part of this work, so please see Chapter 4 for more details.

2.2.5.3. Disruption assay

The same IVT method as described in 2.2.5.2 was used, specifically for co-spotted DNA IVT template arrays containing pairs of RNAs of interest known to interact with each other. Inclusion of a disrupting nucleic acid mimic (NAM) molecule (i.e., PNA) within the IVT mixture therefore sought to explore interaction disruption. This method was developed as part of this work, so please see Chapter 5 for more details.

2.2.6. RNA array probing and visualisation to assess RNA capture on the RNA array

2.2.6.1. RNA Probing Step

In the probing, the immobilised RNA was detected by hybridisation with a short fluorescently labelled single stranded DNA probe specifically designed to base pair to the SA-linker region. A probing mixture, containing the probe, was used to form a 'probing sandwich' between the RNA array slide of captured RNAs and a liftaslip. The probe was diluted in 2 x SSC with 0.1% (w/v) SDS (Appendix 14) to the desired concentration, usually 500 nM. The probing sandwich was then transferred to the humidified chamber and incubated at room temperature protected from light for 30 minutes. At the end of the

probing step, the liftaslip was removed from the RNA array slide, which was then transferred to a Falcon tube for the washing step as follow:

- Submersed in 1 x PBST buffer for 5 minutes
- Extra wash in Milli-Q dH_2O for 30 seconds
- Final dip into Milli-Q dH_2O

After the washing steps the RNA slide was dried in the centrifuge at 1000 rpm for 30 seconds and then visualised. The process is illustrated schematically in Figure 2.8.

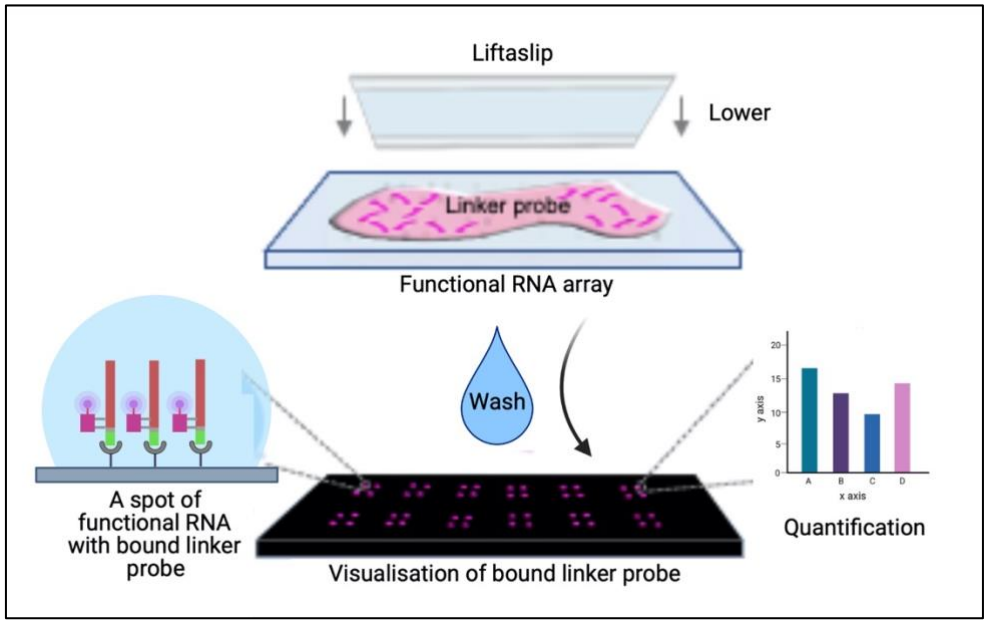


Figure 2.8. Schematic of the RNA array probing and visualisation step to assess RNA capture. A solution of fluorescently labelled SA-linker probe is 'sandwiched' between the RNA array and a liftaslip and incubated. The probe binds to the SA-linker region of the immobilised RNA on the array. Following washing the RNA array is visualised and the fluorescence of the bound-linker probe quantified; indicative of the RNA captured on the RNA array.

2.2.6.2. RNA visualization and data analysis

The probed RNA array from step 2.2.6.1. was scanned using a GenePix 4300A microarray scanner with integrated GenePix Pro 7 image analysis software (Molecular Devices). To detect bound Alexa647-labelled SA-linker probe, the RNA array was visualised

using an excitation wavelength of 635 nm and a Standard Red emission filter. The fluorescence intensity of each spot on the array was quantified using the integrated image analysis software on the microarray scanner.

2.2.7. RNA array probing and visualisation for interaction/disruption assays

2.2.7.1. Probing with fluorescently labelled RNA generated by IVT

Fluorescently labelled RNA was produced by IVT incorporating a Cyanine 3 (Cy3) uridine-5'-triphosphate (UTP) (GE) in the IVT reaction described in section 2.2.3.5. The reaction volume was altered to include a 0.5 mM final concentration of Cy3 UTP (Table 2.8). All fluorescently labelled molecules were protected from light as much as possible for example by using amber tubes.

Table 2.8.
IVT reaction mix for making CY3-labelled RNAs (MegaScript T7 and GE).

Component	Initial []	Final []	Volume (μl)
Nuclease-free dH_2O	-	-	To 20 μL
T7 Reaction Buffer	10 x	1x	2
DNA IVT Template	-	200 ng	-
rNTPs (each - UTP)	10 mM	1 mM	2
Cy3 UTP	5 mM	0.5 mM	2
RNA Pol T7	10 x	1x	2

Purification of the labelled RNA was carried out using the MegaClear (Invitrogen) spin columns as described in the standard protocol and eluted in 15 μL. Samples were run on a 6% (w/v) denaturing gel using gel loading dye from MegaScript T7 Ambion kit and a RiboRuler Low Range RNA Ladder (both from Thermo Fisher Scientific) as a sequence size marker. The concentration of the purified samples was verified using the microvolume UV-VIS NanoDrop 2000 spectrophotometer, and their size and quality confirmed by gel

electrophoresis run on a suitable polyacrylamide gel (method described in section 2.2.7.2). See Appendix 11 for gel of RNA samples for GcvB-Cy3. Probing of the RNA array with the labelled RNA (GcvB-Cy3) was carried out as described for the in 2.2.6, except that 500 nM labelled RNA was used for the probing.

2.2.7.2. GAG-linker probe for detecting GcvB bound to target mRNA

RNA arrays produced following co-spotting experiments potentially include two RNAs bound to the surface following being transcribed at the same time. This usually included a target mRNA surface-immobilised via an appended SA aptamer, and its sRNA binding partner. Detection of the binding partner sRNA involved probing with a fluorescently labelled antisense ssDNA probe specific for the sRNA; in this case an Alexa488-labelled GAG-linker probe designed to bind to a **m**odified linker region of **G**vcB (GcvB-**MG**). Probing of the RNA array with the GcvB GAG-linker probe was carried out as described for the in 1.2.6.

2.2.7.3. Visualisation and analysis

Following a probing step, the RNA slide was visualised using a GenePix 4300A slide scanner (Molecular Devices) using three different wavelengths (488, 532 and 635 nm). For visualising bound GcvB-Cy3, an excitation wavelength of 532 nm and a Standard Green emission filter were used. For visualising bound Alexa488-GAG-linker probe an excitation wavelength of 488 nm and a Standard Blue emission filter were used. The photo multiplier tube (PMT) was adjusted as required to ensure the highest signal possible while preventing signal saturation. The platform integrated software GenePix was used to process the images and quantify fluorescence intensity of the spots.

2.3. Other software used in this thesis

2.3.1. Biorender

Biorender is a widely used web-based tool specifically created for medical and life science illustration, which allows the user to create high quality diagrams. It contains a catalogue of templates and icons commonly used by researchers and life scientists, which have been incorporated as scalable icons in the figures. Biorender has been used to generate all figures in this thesis. Biorender is available at https://app.biorender.com

2.3.2. Microsoft Excel

Microsoft Excel is a software package for analysing spreadsheets of data. It was used throughout this work to analyse data and create graphs of the findings.

2.3.3. PNA design

PNA molecule used in Chapter 5 was designed using the guidelines in the PNA Tool at the PNA Bio website: https://www.pnabio.com/support/PNA_Tool.htm

2.3.4. OligoCalc

Molecular weight for all molecules were calculated using OligoCalc web tool at: https://horizondiscovery.com/en/products/tools/TM-Calculator

Chapter 3

3. *In silico* studies: Identification of novel *trans*-acting sRNAs involved in gene regulation in APP

3.1. Introduction

The strategy used in this project was organised in two parts, specifically into *in silico* studies (this Chapter) and *in vitro* studies; with the *in vitro* studies being subdivided into interaction (Chapter 4) and disruption studies (Chapter 5). Interaction studies heavily rely on accurate sequences since nucleic acids use base-pairing in designated regions of their sequences to bind onto each other.

Although the APP serovar 8 MIDG2331 genome has been previously assembled, its transcriptome has not been publicly characterised, concealing information on RNA features essential to this work, such as the precise sequence boundaries of the predicted novel sRNAs and the untranslated region of their partners mRNAs. Current sRNA prediction tools identify *cis* and *trans*-acting sRNAs alike, but this study focused on *trans*-acting sRNAs, to investigate the versatility of imperfect base-paring interactions between one sRNA and a pool of mRNA genes from a distant locus.

For this reason, the first part of the project focussed on developing a concise *in silico* pipeline to increase prediction accuracy of *trans*-acting sRNAs in APP serovar 8 from a database of computationally predicted novel sRNAs provided by collaborators. The pipeline utilised RNA-seq data and homology studies to narrow the list of sRNA candidates, before proceeding to the prediction of their mRNA partners.

The final output, a single APP serovar 8 novel sRNA candidate and its mRNA partners, was then tested in an experimental interaction and disruption validation pipeline. The sRNA candidate and mRNA target sequences were used to design DNA templates, which were then synthetically produced for the *in vitro* studies, which takes advantage of the novel RNA Array technology, developed within the Callaghan Group at the University

of Portsmouth. The *in vitro studies* of this project are described in Chapters 4 and 5, but a brief description of both the *in silico* and *in vitro* approaches, and how they connect to create the project strategy, is illustrated in Figure 3.1.

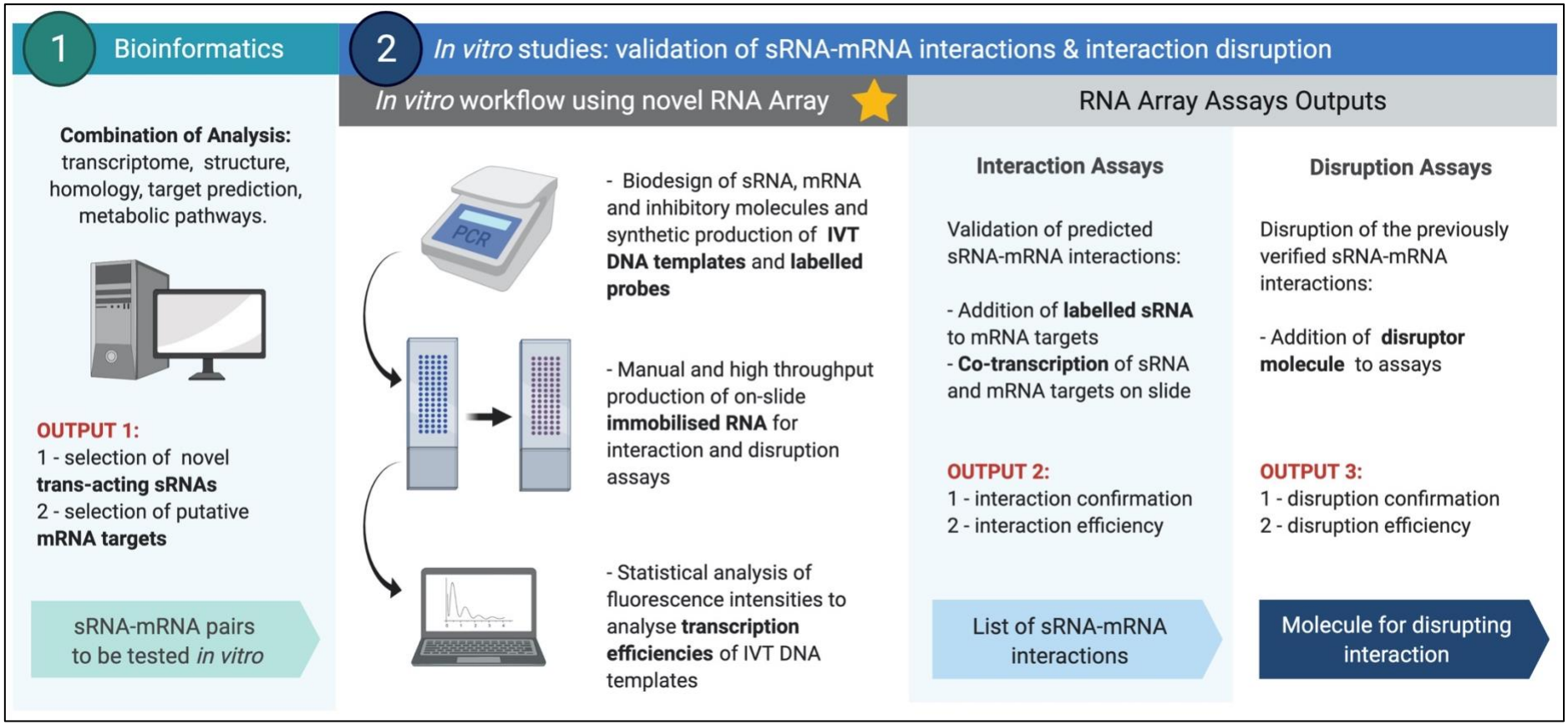


Figure 3.1. Approach to studying novel sRNAs and their putative mRNA partners. 1) The first part of the strategy utilises a combination of bioinformatics methods to identify, predict and analyse novel sRNAs and their putative mRNA targets. The sequences of the selected sRNA-mRNA pairs are used to design and synthetise the molecules probed in the second part of the strategy, including sRNAs, mRNAs and inhibitory molecules. 2) An RNA array method is established to validate the predicted sRNA-mRNA interactions and exploit their disruption *in vitro* using an inhibitory molecule, as a means of proof of concept for the development of a novel antibacterial therapeutic approach.

3.1.1. Novel sRNAs candidates in APP serovar 8

The genome of a clinical isolate APP serovar 8 MIDG2331 was annotated and published in 2016 by collaborators from Imperial College London and University of Viçosa, Brazil (Bossé *et al.*, 2016). The group commissioned the RNA-seq data from the APP species used in this study and provided a comprehensive database of predicted sRNA candidates for MIDG2331, named ICL_UoV_2016 in this manuscript (Appendix 1). Details of RNA-seq data such as full QC report, TPM values and feature counts are reported in Appendix 2.

ICL_UoV_2016 was created using methods established by the collaborators previous studies on APP serovar 5 described in section 1.4 of this manuscript (Rossi *et al*., 2016), and by visual inspection of the RNA-seq data graphic representation in the RNA-seq visualisation software Artemis (Carver *et al*., 2012) by Dr Janine Bossé, a collaborator at Imperial College London. The sRNA candidates added by visual inspection were not identified by the available prediction tools at the time of this analysis, and apart for being expressed in intergenic regions (IGRs) in RNA-seq experiments, these transcripts displayed sRNA characteristics such as promoters, terminators and 3' Poly-U tail.

Computationally predicted sRNAs in APP serovar 8 MIDG2331 were obtained using the strategy from previous studies on APP serovar 5 L20 (Rossi *et al*., 2016) and were named ARRC## (**A**PP **R**egulatory **R**NA **C**andidate and a two-digit number), and the sRNAs identified by RNA-seq data from co-immunoprecipitation of RNAs with Hfq (Co-IP) were labelled RNA## by the collaborators. However, sRNA candidates spotted by RNA-seq visual inspection by Dr Janine Bossé were designated the initials JB followed by a number (JB_##) in this thesis. Both, RNA-seq and sRNA prediction data for MIDG2331 were unpublished during the course of this research.

3.1.2. A pipeline to identify *trans*-acting sRNA candidates and their putative mRNA targets

The ICL_UoV_2016 database was particularly extensive for the purpose of this project, and initially required an analysis to create a reduced list of sRNA candidates and a set of its predicted mRNA targets for interaction testing using the RNA array technology. Thus, a pipeline (Figure 3.2) consisting of computational approaches was created with three specific and sequential objectives; to use APP genomic, and transcriptome data (Figure 3.2 - A) to constrict ICL_UoV_2016 to a refined selection of predicted sRNAs (P-sRNAs) that exhibit structural and regulatory elements present in characterised *trans*-acting sRNAs (Figure 3.2 - B); to select a set of predicted mRNA targets for the chosen P-sRNAs using predictions tools combined with expression information from RNA-seq data (Figure 3.2 - C); to select sRNA candidates whose mRNA partners are involved in metabolic pathways possibly supporting or linked to known virulence factors in APP (Figure 3.2 - D); and finally, to use the output of this pipeline to design and synthetically produce DNA templates required for the RNA array relevant to the selected sRNA and mRNA targets to be investigated in the interaction studies, and to design the inhibitory molecules used in the disruption studies in the second part of this project.

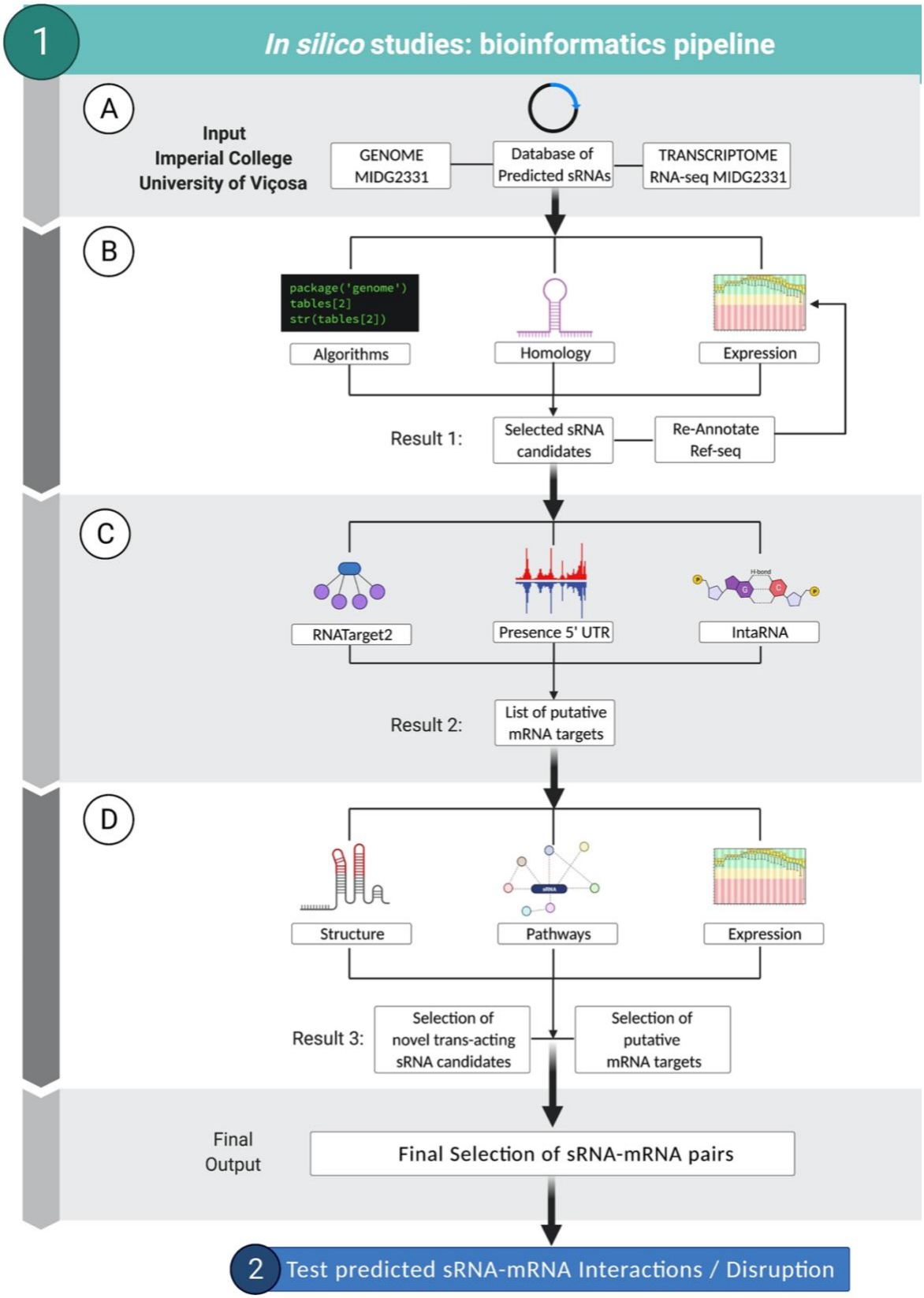


Figure 3.2. Bioinformatics pipeline. A computational approach to identify novel *trans*-acting sRNA candidates and their putative mRNA partners linked to virulence in APP. The pipeline consists of a selection of genomic tools to identify which predicted sRNA-mRNA pairs are applicable candidates to be investigated using the RNA array. The steps and results for each section of the pipeline are describe in the following sections.

3.2. Selecting novel *trans*-acting sRNA candidates in APP serovar 8

The workflow was initiated with the input data from the above-mentioned collaborators (Figure 3.2. - A). The aim of the analysis set in Figure 3.2. – B was to scrutinise the initial database and to select sRNA candidates displaying all the elements established as features of *trans*-acting sRNAs necessary for this research. To achieve this, one of the strategies was to select predicted sRNAs displaying consistent transcriptional features, and to remove those displaying elements common in *cis*-acting sRNAs, such as antisense sRNAs and riboswitches. For this reason, a new set of quantitative and binary parameters was added to the ICL_UoV_2016 database, making it possible to filter them in or out, and select only predicted sRNAs with *trans*-acting characteristics and regular transcriptional signals.

The initial database contained 93 sequences from APP serovar 8 predicted as sRNA candidates, and among them, 22 sequences were computationally identified by Rfam (Nawrocki *et al*., 2014) while 71 were not previously annotated in the MIDG2331 genome. Those that were not already annotated in the MIDG2331 RefSeq file were manually added, and a re-annotated RefSeq draft file (Appendix 2) was created for the purpose of this study only. All 93 sequences showed positive expression levels in wild type and anaerobic growth conditions. However, the final number of predicted sRNA candidates decreased to 27 after selecting for parameters based on sequence conservation and transcriptional signals consistent with *trans*-acting sRNAs in the first section of the pipeline. A further selection for sRNAs identified in Co-IP RNA-seq data showed that only 13 sRNAs were detected. From these, only 9 sRNAs showed full sequence conservation in APP serovar 5 L20.

Biophysical features of each of the final 9 sRNAs were analysed by secondary structure prediction analysis (section 3.2.3), and the draft RefSeq was used to quantify expression levels of these sRNAs in different growth conditions (graphically represented in section 3.2.4). The values for the parameters set for sRNA type, transcription features, expression levels and structural information of these 9 sRNAs was stored in a database named UoP_ICL_UoV_2020 (Appendix 2). Finally, because the RNA array is a new technology, only one putative sRNA candidate was initially selected from the final database to validate the technical strategy and establish the system to create a standardised and universal interaction and disruption validation pipeline to study novel sRNAs.

3.2.1. Annexing parameters to the initial database of novel sRNAs candidates

The ICL_UoV_2016 database already contained binary data related to the expression patterns of its predicted sRNAs (P-sRNAs), and qualitative and quantitative information on transcription signal elements including promoter region, transcription start site (TSS), transcription termination and 3' Poly-U tail, among other essential fields such as transcript ID, name, size, function, locus tag, location in the genome, sequence and experimental validation by the collaborators in previous studies. Table 3.1 below was created to describe the original (black) and added (red) parameters, and to establish the expected value for each parameter used to filter the database. While some parameters depend on prediction tools that use genomic sequence such as OperonMapper (Taboada *et al.*, 2018), others are only identifiable using RNA-seq data combined with R pipelines and visualisation tools. This is the case for the first two parameters in the table, which were used as a guide to detect transcripts during visual inspection by collaborators.

Table 3.1.
List of parameters used to select *trans*-acting novel sRNAs candidates for this study.

Parameter	Description	Expected	Processing Software
1 RNA-seq	P-sRNA expressed in any RNA-seq experiment	Yes	*JBrowse, Bedtools, Bioconductor*
2 Co-IP	P-sRNA expressed in Co-IP Hfq flag experiments	Yes*	*JBrowse, Bedtools, Bioconductor*
3 ORF	P-sRNA overlaps ORF on opposite strand	No	*JBrowse, R script, ORFfinder*
4 Operon	P-sRNA locus within predicted operon	No	*JBrowse, R script, OperonMapper*
5 UTR	P-sRNA overlaps UTR on same strand	No	*JBrowse, Bedtools, R script*
6 Promoter	Presence of promoters at (-10) or (-35)	Yes	*PePPer, ICL_UoV_2016, R script*
7 TSS	Clear cliff at start site in TSS-enriched experiments	Yes	*JBrowse, Rockhopper, R script*
8 Terminator	Presence of Rho-independent transcription terminators	Yes*	*TransTerm, RNAFold*
9 Poly(U)	Presence of 3' Poly(U) tail	Yes	*JBrowse, Rockhopper, R script*
10 Homology	Conservation in other APP strains and bacterial species	Yes*	*BLASTn, GenBank, Jalview*

* Non mandatory expected value for selection of novel sRNA candidates

The table contains a column for the parameter measured (newly added parameters in red), a description of that parameter, an expected binary value (yes or no) to filter the database and increase *trans*-acting sRNA prediction accuracy, and the software used to process the data from RNA-seq or APP serovar 8 sequences.

Parameter 1 (RNA-seq) unconditionally required that P-sRNAs in the original database ICL_UoV_2016 were expressed within intergenic regions (IGRs) of the annotated genome in at least one growth condition of the RNA-seq experiments. This allowed P-sRNAs predicted by computational approaches presenting no expression to be negatively flagged

and removed from the list. All 93 P-sRNAs were expressed in both, wild type and anaerobic conditions. Parameter 2 (Co-IP) disclosed whether a P-sRNA is present in the Hfq Flag (Co-IP) experiments, using featureCounts from R Bioconductor. P-sRNAs showing reads >10 (arbitrary minimum) had their expected value set to a non-mandatory binary positive. This is because a great number of known sRNAs show no interaction with Hfq (Livny, 2012), meaning that a portion of sRNA candidates would be excluded if this parameter is set to a mandatory positive, where only Hfq binding P-sRNAs would be considered.

The values of parameters 3 to 5 are based on RNA-seq data combined with ORF and operon prediction tools. These parameters were added to the original database, and although they cannot accurately predict the category of a sRNA on their own, the expression patterns that they represent may resemble those of *cis*-acting sRNAs or patterns that are exceptionally inconclusive to be considered in this study. A general overview of how expression patterns of *cis* and *trans*-acting may be graphically represented is illustrated in Figure 3.3.

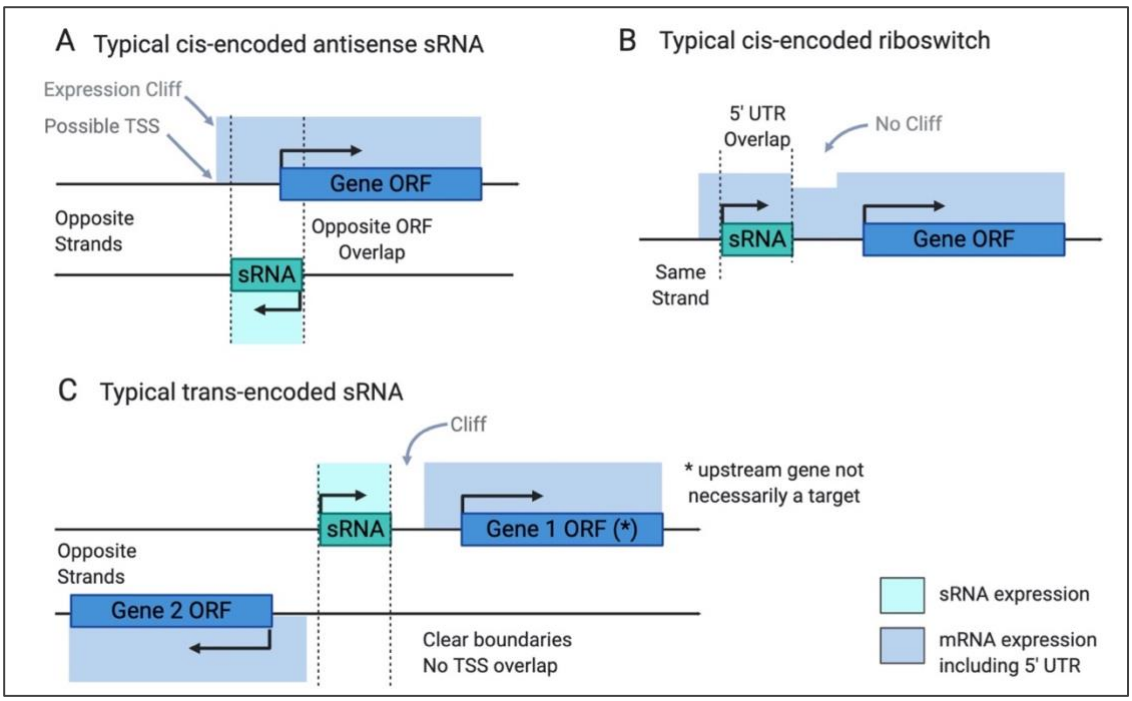


Figure 3.3. Expression patterns of *cis* and *trans*-encoding sRNAs. A variety of visualisation software and UTR finding scripts use transcript reads from RNA-seq data to represent expression boundaries. In this figure, transcript expression is represented in blue and green shaded areas and the boundaries may indicate signals such as a transcription start site (TSS). **A)** Genes for *cis*-encoded sRNAs are usually located in the opposite strand of their target gene, overlapping its 5' UTR and ORF. **B)** Riboswitches are situated on the 5' UTR of the gene that they regulate. **C)** *Trans*-encoded sRNA genes occupy a different location from their targets and although their expression pattern may conflict with an unrelated gene, transcription signals such as transcription start site and Poly-U are separated from neighbouring genes (Li *et al.*,2012).

Parameter 3 (ORF) enquires whether a P-sRNA expression overlaps the ORF of a gene on the opposite strand, a characteristic of *cis*-acting antisense sRNAs, and because the aim is to identify *trans*-acting sRNAs this value was set to mandatory negative. Some parameters were set to determine if the expression pattern of a P-sRNA is noticeably conflicting to be considered. This may happen because the P-sRNA is located within an operon (parameter 4), where expression boundaries of the transcripts are indistinguishable between cistronic UTR and P-sRNA. Parameter 5 (UTR) selects for P-sRNAs with expression patterns that overlap the 5' UTR of a gene on the same strand, a trait observed in riboswitches. However, some P-sRNAs may be located downstream the 5' UTR of a gene and display undeniable independent start of transcription (TSS) and expression pattern. For the reasons presented, values for parameter 4 to 5 were set to mandatory negative.

Parameters 6 to 10 contained qualitative and quantitative information in ICL_UoV_2016 which were converted to binary values to represent the presence of transcription signals (Promoter, TSS, Terminator and Poly-U), and sequence conservation in other APP strains and species (Homology). Although most information was created by collaborators, these values were updated in this research. Promoter and Rho-independent terminator predictions remained the same, but the terminator parameter was set as non-mandatory, as some P-sRNAs may present a stem loop at the 3' end not detected by TransTerm. A poly-uracil tail (Poly-U) is a sequence of a minimum of three uracil residues that follow a stable stem-loop at the 3' end and is considered an important regulatory signal for transcription termination (Wilson & Von Hippel, 1995), and thus considered mandatory in this work.

Values for TSS included extra manual curation of locally enriched transcript expression by identifying regions where frequency of reads dropped or surged drastically. Graphically, these regions are similar to expression cliffs illustrated in Figure 3.3 and their patterns are calculated using BAM files generated from RNA-seq experiments. Finally, the homology parameter was set to mandatory conservation to filter the P-sRNA list, as *trans*-acting sRNAs are highly conserved across the different strains of APP. More specifically, only P-sRNAs showing > 90% sequence similarity to the APP serovar 5 L20 strain were selected for further studies, because mRNA target prediction tools such as TargetRNA2 and other genomic software used in this study consistently contain the genome for this strain.

3.2.2. Selecting plausible sRNAs candidates presenting *trans*-encoding features

The parameters were organised in a graph (Figure 3.4), which shows the ratio of the binary values for each, and the decision flow taken to narrow the initial database, resulting in a reduced selection of 9 predicted novel *trans*-acting sRNA candidates. All 9 P-sRNA candidates presented the expected values set in Table 3.1, including RNA01, RNA03, RNA05 (computationally predicted ARRC01 in APP serovar 5 L20), RNA16, RNA17, RNA19, RNA20, JB_04 and JB_39, which were further analysed for insights on predicted secondary structure regulatory elements and transcript expression levels in the next sections.

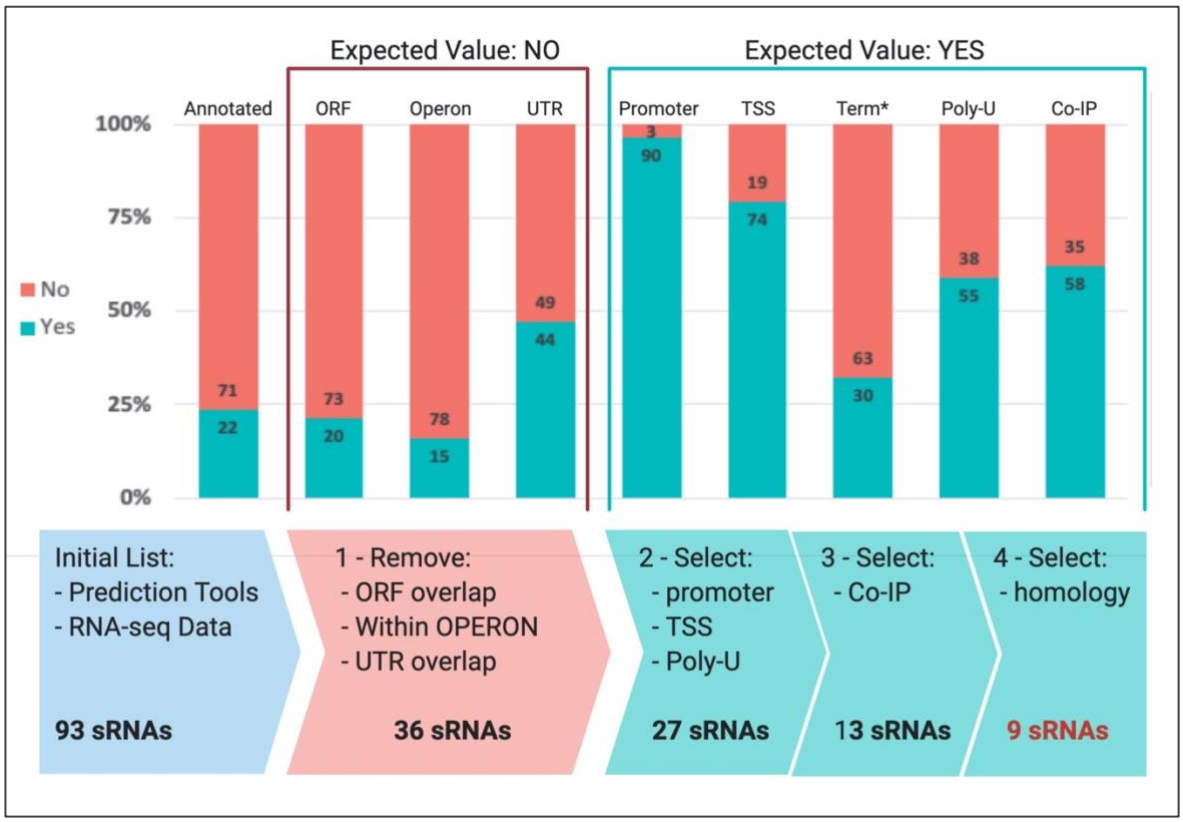


Figure 3.4. Distribution of binary values by parameter and decision flow. The database started with 93 P-sRNAs detected by prediction tools and RNA-seq. All P-sRNAs displaying *cis*-encoding features were removed from the list in step 1, and step 2 filtered for P-sRNAs presenting transcriptional elements, leaving 27 candidates. Although Hfq binding and sequence conservation parameters were not a mandatory characteristic of *trans*-acting sRNAs, these parameters were relevant to lessen the number of candidates for this study in step 3 and 4, thus only candidates with positive values were included in the final list of 9 predicted novel sRNAs. Parameters marked with (*) were not used to refine the database.

3.2.3. Structural analysis of the selected sRNA candidates

When investigating RNA-RNA interactions, the structures established during and after RNA transcription deserves special attention, since they are directly involved in the biological processes regulated by non-coding sRNAs. Although the geometry associated with these structures gives clues about the interplay between these molecules, including sRNA-mRNA interactions, the RNA interactome is generally not well characterised (reviewed in Zampetaki *et al.*, 2018). Palindromic sequences are a common feature found in sRNAs. These sequences form double stranded regions arranged in stem-loops along the sRNA, making it more structurally compact and less prone to degradation by RNases (Zundel *et al*., 2009; reviewed in Deutscher, 2015). Secondary structure prediction (Figure 3.5) by RNAFold (Gruber *et al.*, 2008) shows that all of the 9 P-sRNAs may form at least 2 stem-loops and, apart from JB_04, display a stem-loop at the 3' end, with high base-pairing probability (indicated with red arrows in Figure 3.5), demonstrating the presence of a possible transcriptional termination signal. The predicted Gibbs free energy change (ΔG) showed a negative value for all predicted secondary structures in RNAFold, and RNA05/ARRC01 and JB_39 revealed the lowest values suggesting the most stable structure in comparison with the other sRNAs.

The segments of sequence in sRNAs that interact with mRNA targets through imperfect base-paring are usually located in single strand regions and referred as seed regions or simply interaction regions. This element is computationally predictable with a variety of tools including TargetRNA2 and IntaRNA. Target prediction of the P-sRNAs by TargetRNA2 identified preferable interaction regions in all of the 9 P-sRNAs, which are indicated with green arrows in Figure 3.5. However, the sRNAs RNA03 and RNA16 revealed their seed region in double stranded areas of predicted stem-loops with low base-pair probability. Although analysis of structural elements is based on predictions rather than experimental results, it offers important insights to support an informed selection of an sRNA candidate. P-sRNAs with a predicted stable 3' stem-loop and with its preferred seed region located on single stranded areas offer a better justification for their selection.

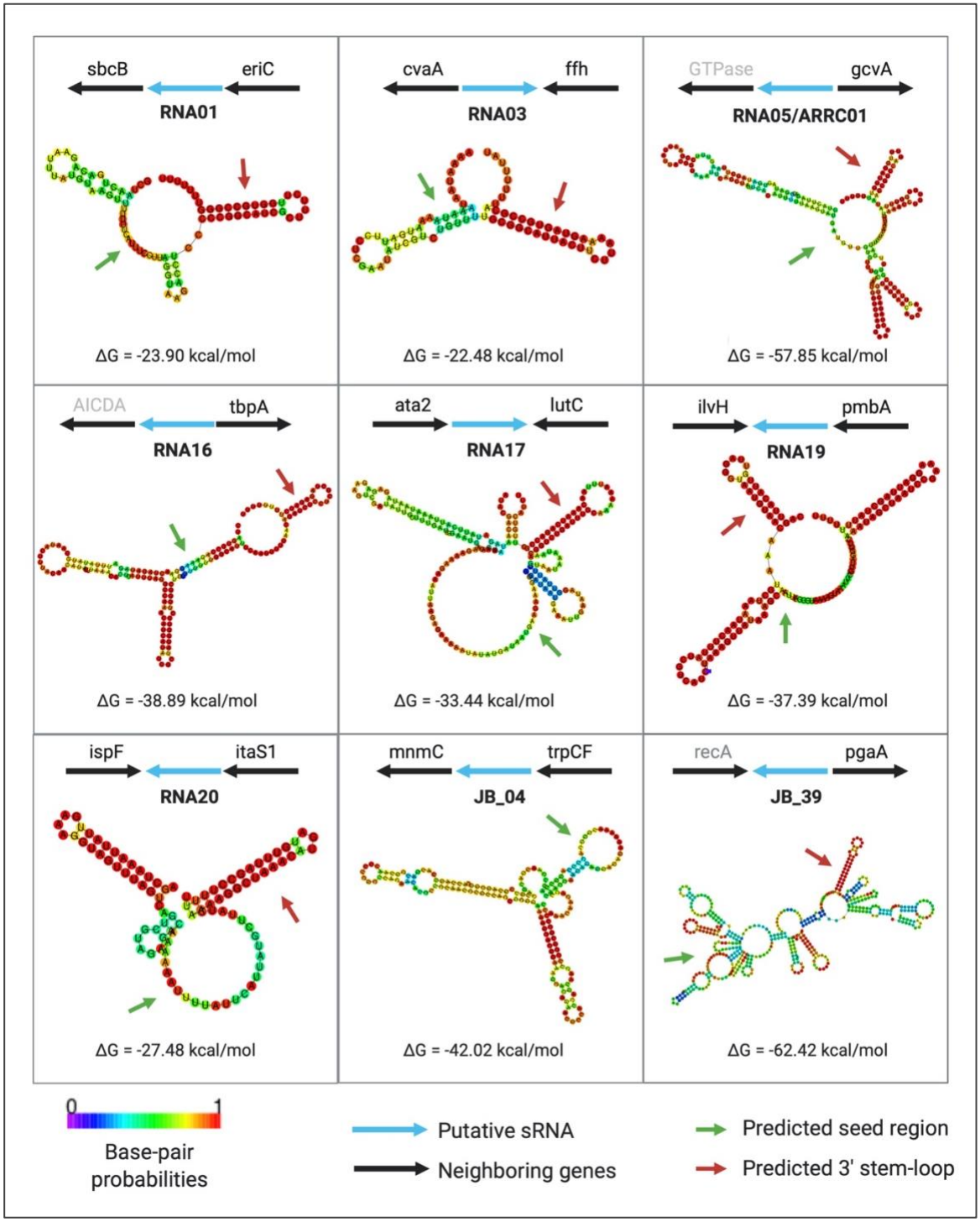
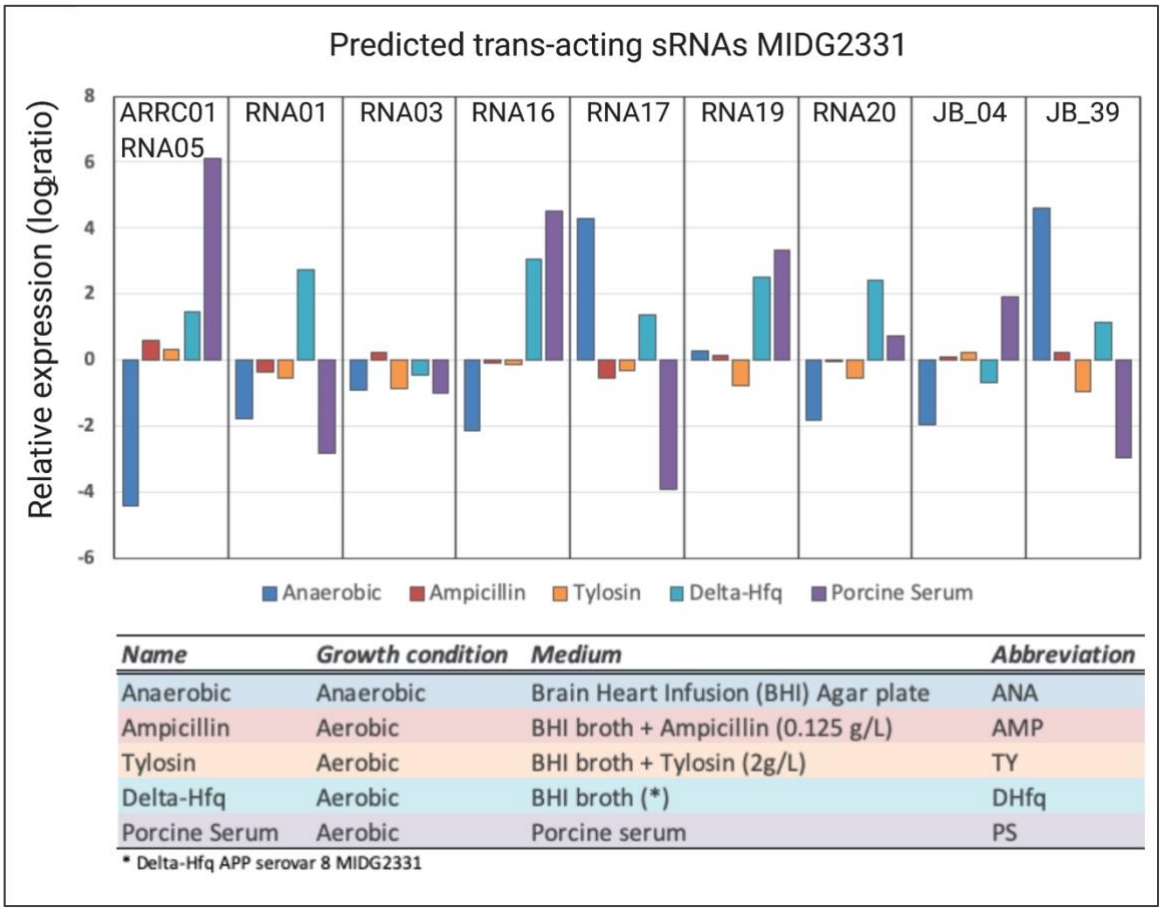


Figure 3.5. Structural analysis of *trans*-acting sRNA candidates. The predicted secondary structures of the selected sRNA candidates by RNAFold forecast the formation of stem-loops in all of them. A red arrow indicates the presence of a 3' end stem-loop with a high base-pairing probability in a coloured gradient scale ($0 < p > 1$), where a pair of sites with 100% canonical pairs is coloured dark red. A green arrow indicates the TargetRNA2-identified preferable target interaction region. All structures present a negative ΔG representing spontaneous formation at 37°C. Neighbouring genes are represented in black arrows which indicates their location in either, forward or reverse strand in relation to the sRNA. Finally, the predicted *trans*-acting sRNAs are illustrated with a blue arrow, also indicating their location and transcription direction.

3.2.4. RefSeq re-annotation and expression analysis of the selected sRNAs

The sequence information of the selected sRNAs such as gene name and boundaries, were manually added to the RefSeq file MIDG2331 and saved as a draft. This made it possible to quantify the expression levels of these putative sRNAs, which were previously invisible to expression quantification tools such as featureCounts (Liao *et al.*, 2014) in Bioconductor (Gentleman *et al.*, 2004) and BEDtools (Quinlan & Hall, 2010), which counts gene reads generated by RNA-seq experiments that are mapped to the associated ORF and non-coding RNA (ncRNA) features. RNA-seq experiments were commissioned and planned by collaborators (details on methods section 2.1). Because the main goal of these experiments was to investigate novel sRNAs in APP serovar 8 MIDG2331, the technical repeats for each growth condition were pooled in one sample prior to sequencing. This is a common procedure utilised to reduce costs and optimise statistical outputs by obtaining a higher RNA quantity, which aids detection of smaller transcript sequences, and by averaging the transcript levels across samples (Auer & Doerge, 2010; Assefa *et al.*, 2020). However, sample pooling presents risk of bias, as it is not possible to verify the exact total amount of RNA from each pooled sample or whether they contributed equally to the final pool (Rajkumar *et al.*, 2015; Conesa *et al.*, 2016).

Another obstacle is related to differential expression analysis pipelines, which usually require technical repeats to present a more accurate picture to establish expression levels. For this reason and for the purpose of this study only, RNA-seq analysis focussed on two factors from RNA-seq data; first to verify whether these putative sRNAs were expressed by examining their transcript count reads when APP serovar 8 MIDG2331 is cultivated in aerobic and anaerobic growth conditions, and from MIDG2331 Hfq Co-IP experiments. Second, to analyse the expression ratio of the predicted sRNA transcripts of MIDG2331 grown in a specific condition in comparison to MIDG2331 aerobically grown in **B**rain, **H**eart **I**nfusion (BHI) broth or BHI Agar plate media, named as wild type (WT for reference in this thesis). Expression ratio was normalised (log_2ratio) and graphically represented in Figure 3.6.



Name	*Growth condition*	*Medium*	*Abbreviation*
Anaerobic	Anaerobic	Brain Heart Infusion (BHI) Agar plate	ANA
Ampicillin	Aerobic	BHI broth + Ampicillin (0.125 g/L)	AMP
Tylosin	Aerobic	BHI broth + Tylosin (2g/L)	TY
Delta-Hfq	Aerobic	BHI broth (*)	DHfq
Porcine Serum	Aerobic	Porcine serum	PS

* Delta-Hfq APP serovar 8 MIDG2331

Figure 3.6. Relative transcript expression of the predicted *trans*-acting sRNAs. Transcript levels of the selected sRNA candidates were quantified in all growth conditions considered in this study (Table 3.1), and their transcript yield relative to their WT in broth or Agar plate media was calculated and normalised ($Ratio_2log$) for all 9 P-sRNAs. Expression levels of MIDG2331 wild type (WT) in **anaerobic** (ANA) BHI Agar plate was compared against WT in **aerobic** growth condition in Agar plate (blue bars). MIDG2331 WT with added **AMP** (red bars) and **TY** (orange bars) experiments were carried out in BHI broth in aerobic growth conditions and both compared to MIDG2331 WT grown in aerobic BHI broth. The yield of MIDG2331 Hfq knockout (**DHfq**) in BHI broth (green bars) was compared against MIDG2331 WT grown in BHI broth. Finally, transcript ratio of MIDG2331 WT grown in porcine serum (PS) was compared against MIDG2331 WT grown in aerobic BHI broth (purple bars).

Bacterial RNA-seq experiments benefit from different growth and media conditions which simulate, as close as possible, the different environments in which pathogenic bacteria survive and thrive in the host to become virulent. APP is a facultative anaerobic pathogen, and RNA-seq experiments carried out in anaerobic growth conditions offers insights on which sRNA transcripts are present when APP cells are free of molecular oxygen in relation to aerobic environments. The results showed that compared to aerobic growth conditions, most P-sRNAs have negative or very low transcript yield when MIDG2331 WT is grown in **anaerobic** conditions, except for RNA19 and JB_39 (> +4.0-fold).

Many *trans*-acting sRNAs interact with the facilitator protein Hfq, that may protect sRNAs from ribonucleases cleavage. However, Hfq may induce cleavage of some sRNAs when paired with their mRNA partners (Massé *et al.*, 2003). Thus, MIDG2331 Hfq-knockout experiments were undertaken to investigate how the number of sRNA transcripts may be altered without the presence of Hfq. **Delta-Hfq** MIDG2331 grown in aerobic BHI broth shows a positive fold in transcript reads for the predicted sRNAs, with RNA01, RNA16, RNA19 and RNA20 displaying an even higher yield (~> +2.0-fold).

When grown in selective media, with added antibiotics **ampicillin** (AMP) and **tylosin** (TY), all 9 P-sRNAs have a low transcript ratio output (~< ±0.5-fold). This result may suggest a lower total RNA extraction from AMP and TY samples due to antibiotic activity affecting bacterial cell growth, or/and lower expression/high degradation of these P-sRNA transcripts. In comparison to added antibiotic media, a much higher change in transcript yield (~> ±2.0-fold) appears when MIDG2331 WT is grown in aerobic **porcine serum** media, with exception of RNA03 and RNA20 (~< 0.5-fold). The choice of porcine serum media for RNA-seq experiments is an attempt to simulate the physiological conditions from host cells from pig respiratory organs, especially regarding levels of nutrient availability during virulence stages. The P-sRNA RNA05/ARRC01 displayed the higher transcript fold (> +6.0-fold) in porcine serum growth media in comparison to all other sRNAs in any condition. Except for RNA01, RNA03 and RNA19, sRNA transcript yields in anaerobic and porcine serum display opposite fold directions.

3.2.5. A refined table of *trans*-acting sRNA candidates for APP serovar 8

Once the predicted sRNAs had been scrutinised, information on transcript expression levels and secondary structure prediction was added to the final database UoP_ICL_UoV_2020, and a table with the 9 selected putative sRNAs for APP MIDG2331 was created (Table 3.2 below). The bioinformatics pipeline part of the project strategy was designed to serve the interaction and validation pipeline for the RNA array by selecting only the most viable predicted *trans*-acting sRNAs. All the 9 P-sRNAs presented the characteristics desired for this project.

Table 3.2.
List of putative *trans*-acting sRNAs in APP serovar 8 MIDG2331

ID	Annotation	Genome Position	Length	Strand	Promoter (-10)	Promoter (-35)	Terminator	3' - Poly(U)	Upstream gene	Downstream gene	Prediction *in silico*	Northern blot	RNA-seq Highest Yield
ARRC01	MIDG2331_00136	149359..149554	196	-	TATAAT	TTGTCA	--	7	gcvA	MIDG2331_00135	Yes	220 (p21)	Porcine S
RNA01	na	738604..738689	86	-	TCTTAA	GAACTT	term_471	7	eriC	sbcB	nd	85 (p64)	Delta-Hfq
RNA03	na	2035731..2035807	77	+	TATCAT	TTCGCA	term_1336	6	tRNA-Asn(gtt)	ffh	nd	--	Ampicillin
RNA16	na	195041..195194	153	-	TATACT	TTGCTT	--	8	tbpA	MIDG2331_00179	nd	--	Porcine S
RNA17	na	516723..516916	194	+	TATAAT	ATGCTC	term_330, te	3	ata_2	lutC	nd	--	Anaerobic
RNA19	na	808647..808776	130	-	TAGAAT	TTGACA	term_516	7	ilvH	pmbA	nd	--	Porcine S
RNA20	na	896366..896461	101	-	TAGAAT	TTGCAA	term_569, te	3	itaS1	ispF	nd	--	Delta-Hfq
JB_04	na	968916..969087	172	-	TAGACT	TTCTCA	term_622	4	trpCF	mnmC	nd	--	Porcine S
JB_39	na	2218790..2219245	456	-	TCTTA	GAACTT	--	4	MIDG2331_02148	pgaA	nd	--	Anaerobic

(na) Sequences not previously annotated in *A. pleuropneumoniae* serovar 8 MIDG2331.
(nd) Sequences not detected in *in silico* methods.

3.2.6. RNA05/ARRC01 is the global sRNA regulator GcvB

Sequence identity studies of the selected putative sRNAs have been previously inferred by collaborators by homology search using BLASTn (Johnson *et al.*, 2008). The sequence of RNA05/ARRC01 is homologous to the sRNA GcvB, a widely characterised *trans*-acting sRNA in *E. coli* and *Salmonella* (Urbanowski et al, 2000; Sharma et al, 2007; Miyakoshi, 2015). However, the 8 remaining sRNAs are novel *trans*-acting sRNA candidates, not previously identified in other species. From this point in the thesis, the sequence for RNA05/ARRC01 is referred as the putative sRNA GcvB in APP serovar 8 MIDG2331 or simply APP8 GcvB for short.

Genomic comparative studies are relevant to identify whether known interaction regions and structural conformation of an sRNA are conserved within and across the species, especially as the APP8 GcvB is yet to be characterised, thus this was carried out using APP8 GcvB DNA nucleotide sequence extracted from this study. The first homology investigation compared APP8 GcvB against a selected set of well characterised virulent bacteria (Appendix 3), including pathogenically relevant species within and outside the *Pasteurellaceae* family to study conservation along bacterial phylogeny. The sequences were then aligned using the web tool ClustalW (Goujon *et al*., 2010) and visualised and labelled with the program JalView (Waterhouse *et al*., 2009) as detailed in Figure 3.7. Residues matching 100% identity threshold were coloured purple in an attempt to identify sequence conservation shared by all species in the alignment.

The two most significant conservation regions, labelled as CV1 (APP8 GcvB position 78-90, sequence 5’ – UGUUGUGUUUGCA – 3’) and CV2 (APP8 GcvB position 137-146, sequence 5’ – ACUUCCUGUA – 3’), reveals a G/U-rich region characterised in previous studies of GcvB for *Salmonella* (Sharma *et al.*, 2007) and other sRNAs (reviewed in Updegrove *et al.*, 2015) that targets C/A-rich regions in the sequence of their mRNA targets. The alignment reveals that although APP8 GcvB has a low sequence similarity across Protobacteria (average < 40%), the highly conserved regions shared with these species may elucidate sRNA functional elements in the putative APP8 GcvB.

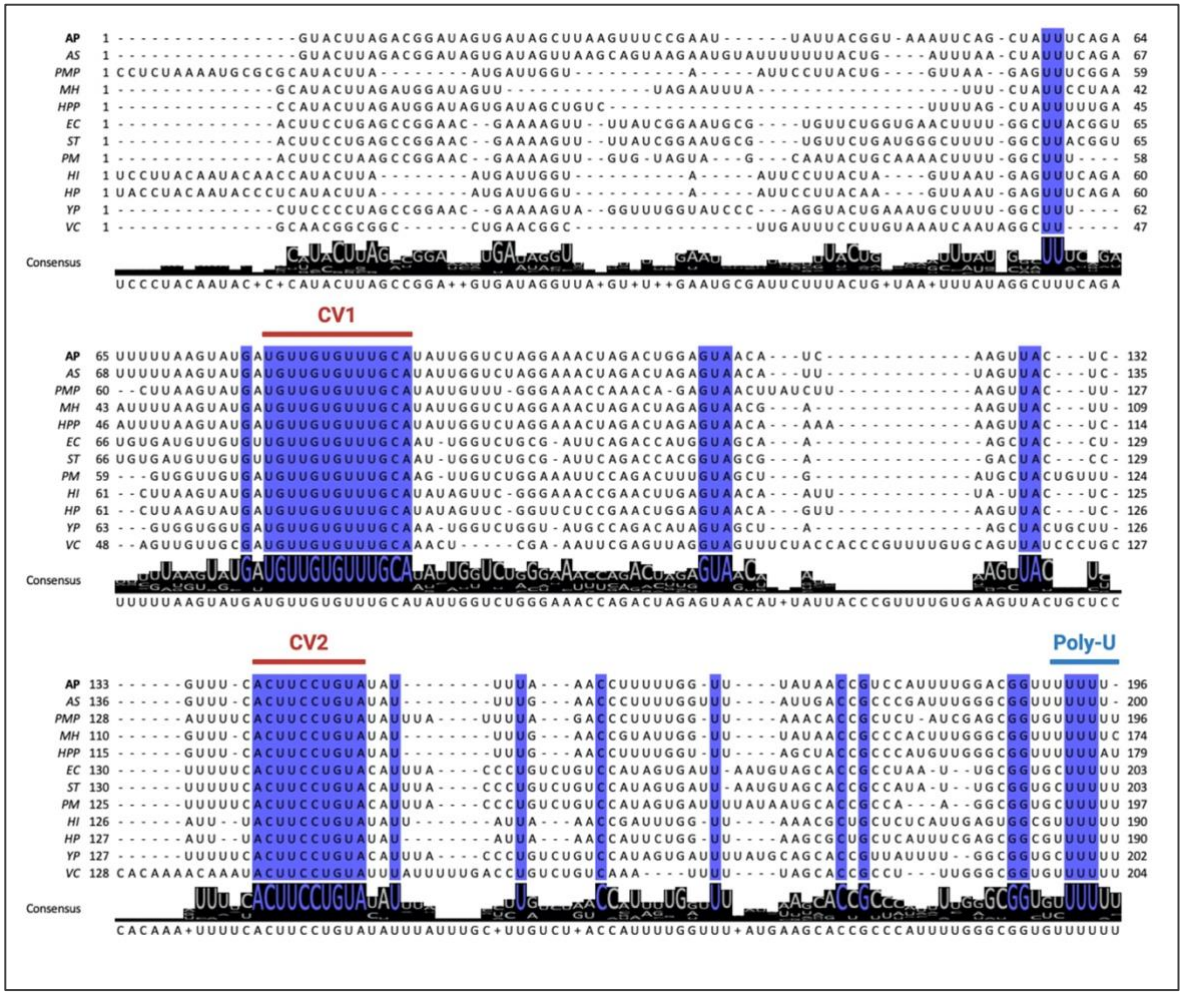


Figure 3.7. Sequence alignment of GcvB across Gram-negative species. The putative APP8 GcvB was aligned with other virulent bacterial species including members of the *Pasteurellaceae* family. The species selected were abbreviated as the following: **AP:** *Actinobacillus pleuropneumonia*e, **AS:** *Actinobacillus suis*, **PMP:** *Pasteurella multocida*, **MH:** *Mannheimia haemolytica*, **HPP:** *Haemophilus parasuis*, **EC:** *Escherichia coli*, **ST:** *Salmonella enterica*, **HI:** *Haemophilus influenza*, **HP:** *Haemophilus parainfluenzae*, **PM:** *Proteus mirabilis*, **YP:** *Yersinia pestis*, and **VC:** *Vibrio Cholerae*. Residues with 100% similarity identity were coloured purple and their consensus highlighted at the bottom of the alignments. Two significant conservation regions were labelled in red as CV1 (AP position 79-91) and CV2 (AP position 138-147), and a third region covering a repetition of uracil (Poly-U) at the 3' end was highlighted in blue.

All species display a varied size of Poly-U at the 3' end, indicating that other structural elements may be conserved in APP8 GcvB, but not revealed in the alignment such as palindromic sequences forming stem-loops. For this reason, the characterised *Salmonella* GcvB was selected from the list of Gram-negative species to investigate whether these two conserved regions in APP8 GcvB (CV1 and CV2) were located in similar regulatory regions of *Salmonella* GcvB, and to investigate the location of predicted stem-loops. The secondary structure of GcvB was predicted using RNAFold for both, APP8 and

Salmonella, and annotation of structural elements of GcvB in *Salmonella*, proposed by Sharma and colleagues (2007), were labelled in Figure 3.8.

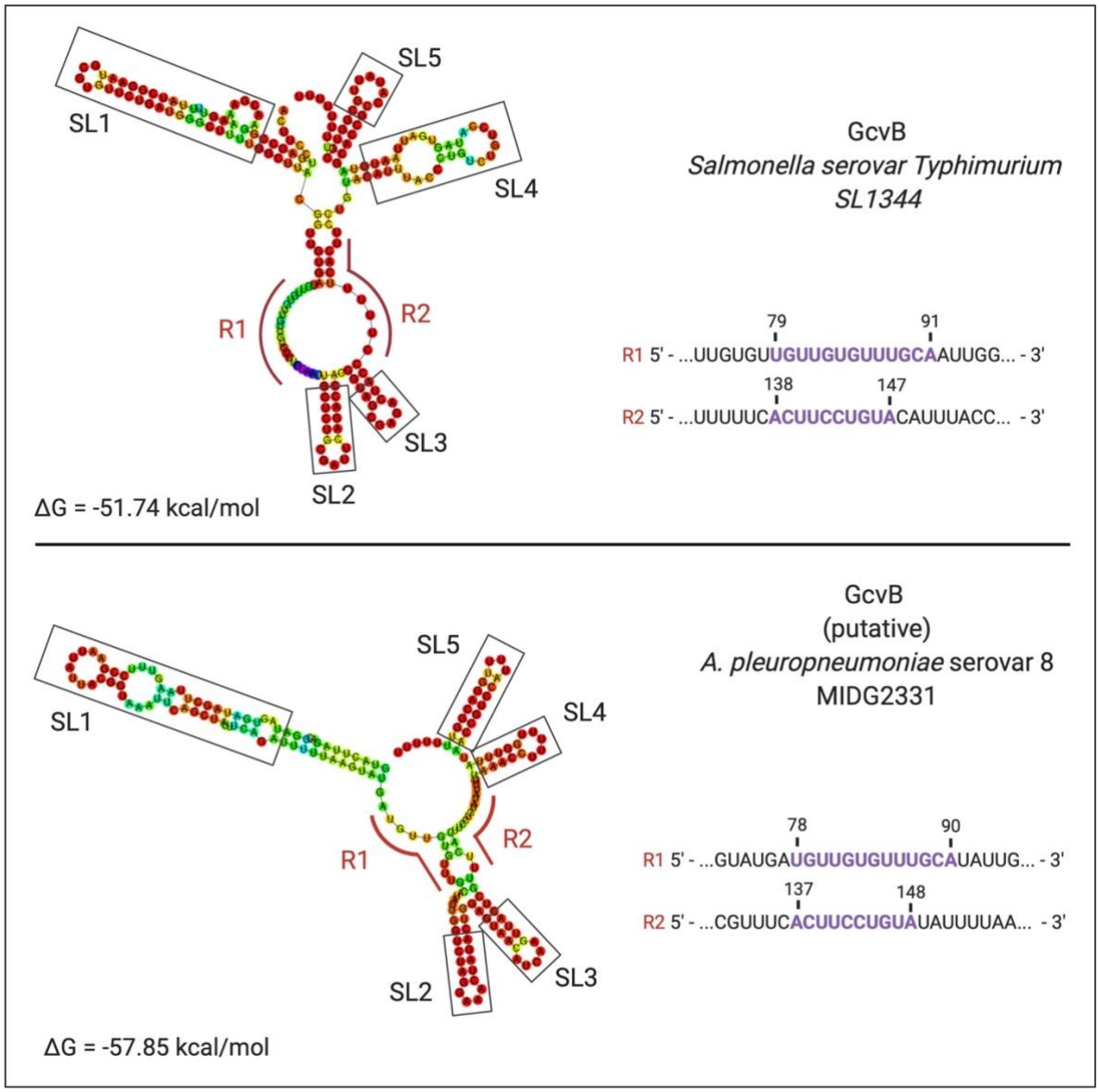


Figure 3.8. Predicted structural comparison of APP8 GcvB and *Salmonella* GcvB. Predicted stem-loops were highlighted by grey boxes and named SL 1 to 5. The two proposed interaction regions of *Salmonella* GcvB were indicated with a red line and labelled R1 and R2. These locations were matched in the APP8 GcvB to represent positional homology. The conserved sequences CV1 and CV2 were highlighted in bold purple and its location matched R1 and R2 for both sRNAs.

Although the nucleotide sequences of these sRNAs have a similarity of around 60% only, both predicted structures presented folding similarities such as the presence of 5 stem-loops (SL1 to SL5), each in approximately corresponding locations along the sRNA, including the termination region. The conserved regions in the alignment for *Salmonella*,

position 79-91 and position 138-147, consecutively overlapped regions that interact with target mRNAs (R1 and R2) as revealed in a study in *Salmonella* (Sharma *et al.*, 2011). When compared to APP8 GcvB, R1 and R2 have approximate positions in the sRNA, which reveals a partially single stranded area, where the conserved regions overlap at 78-90 for R1 and 137-148 for R2 in APP8 GcvB. Predicted structural comparison with *Salmonella* reveals that APP8 GcvB may hold functional elements that are conversed in well-characterised Protobacteria species and across the *Pasteurellaceae* family.

The second alignment analysed how the sequence of APP8 GcvB is conserved across a selected set of APP strains (Appendix 4), and the output (Figure 3.9) revealed 100% conservation of key predicted functional elements such as G/U-rich sequences in the predicted interaction regions (R1 and R2) and a 3' Poly-U of 6 uracil residues. The only discrepancies displayed in the compared serovars involved a two-nucleotide mismatch at positions 35 and 36 among serovars 1, 9 and 12, and a one-nucleotide disparity at position 154 among all serovars except serovar 2. The alignment revealed that the putative GcvB is highly conserved amongst APP strains and the mismatches are unlikely to affect GcvB structure.

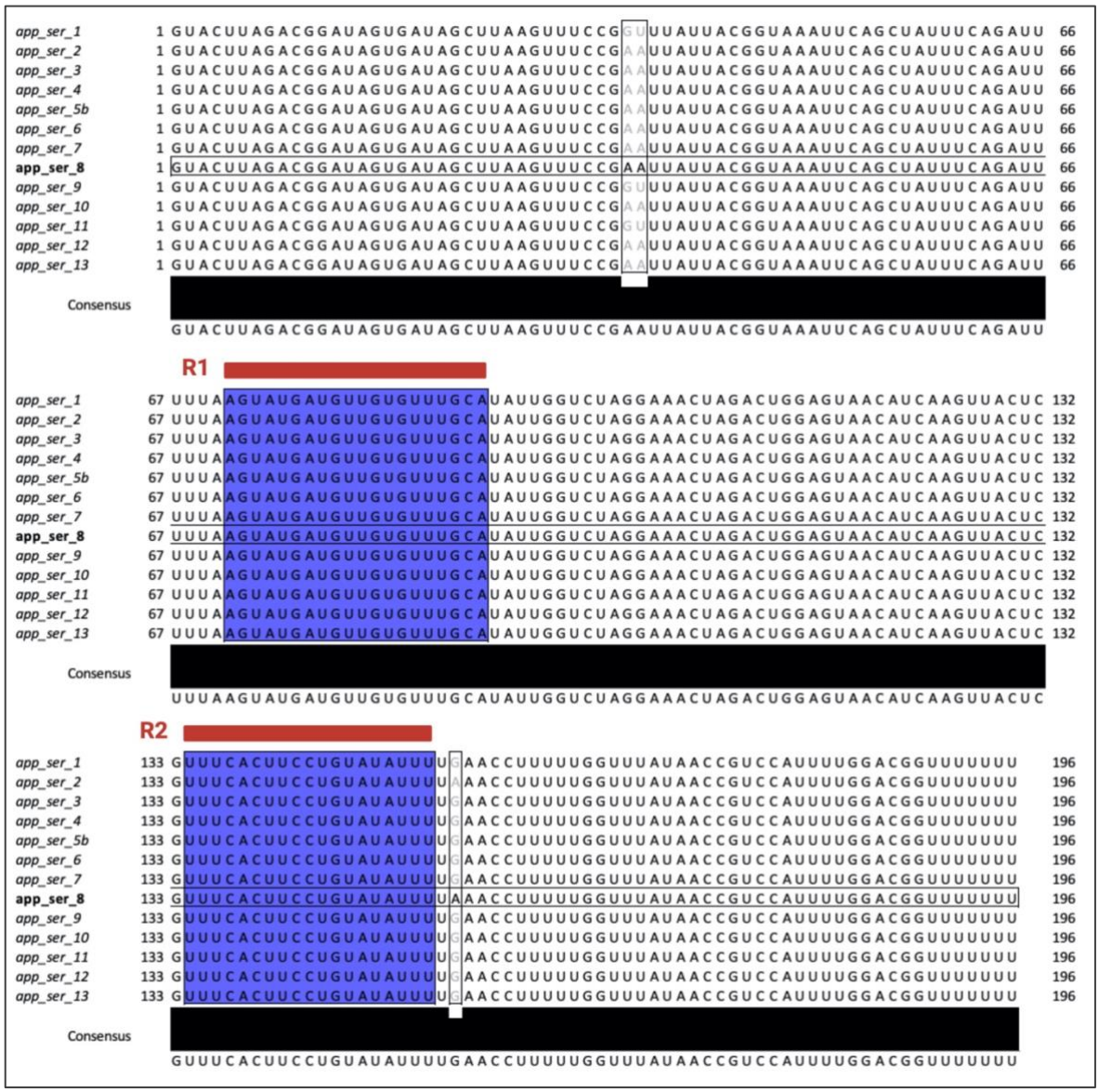


Figure 3.9. GcvB among different APP serovars. Alignments were performed using MIDG2331 GcvB sequence as a reference against a selection of 13 annotated strains of APP (1-4074, 2-4226, 3-JL03, 4-M62, 5b-L20, 6-Femo, 7-AP76, 9-CVJ13261, 10-D13038, 11-56153, 12-1096 and 13-N273, accession numbers in Appendix 4) using the software JalView. The predicted interaction regions R1 and R2 are indicated in red, similarity identity coloured purple for 100% and mismatches shaded in grey.

3.2.7. Selecting the best sRNA candidate to set up the RNA array pipeline

From all the sequences from the predicted sRNAs in Table 3.2, the only predicted sRNA presenting a homologous sequence to a well characterised *trans*-acting sRNAs was RNA05/ARRC01, the putative APP8 GcvB. The sequences relevant to the remaining predicted sRNAs were not previously annotated in other species, making them the perfect candidates for the RNA array validation pipeline. However, because the RNA array is a novel biotechnology, only one set of sRNA and its mRNA pairs was initially required to set up the system and create a standardised RNA array methodology for use as a pipeline to consistently predict, validate and disrupt sRNA-mRNA interactions of novel bacterial sRNAs.

Thus, APP8 GcvB is the best initial candidate to set up the RNA array as it is not characterised in APP, but its sequence homology studies show that functional features are conserved in well characterised *trans*-acting sRNAs. This offers two advantages, including higher probability of predicting true mRNA partners, based on its predicted structural features such as target interaction region located in single strand area of the sRNA, and a better change of synthetically producing a stable molecule due to the presence of various palindromic sequences possibly forming structured stem-loops. Besides, the conservation of G/U-rich sequences are extremely important to predict mRNA partners in programs such as TargetRNA2 and IntaRNA, that uses structural profiling information for characterising sRNA-mRNA interactions rather than base-paring probability alone. For this reason, APP8 GcvB was taken forward to the next set of analysis of mRNA target predictions.

3.3. Prediction of mRNA targets of putative sRNA GcvB in APP

3.3.1. Using TargetRNA2 to predict mRNA targets

A search for potential mRNA targets of the putative sRNA GcvB in APP serovar 8 MIDG2331 (APP8 GcvB) was performed using the program TargetRNA2 (Kery *et al.*, 2014), which identifies the putative sRNA interaction regions in the target mRNA 5' untranslated region (5' UTR). However, genome MIDG2331 used in this study was not available in the TargetRNA2 databases during the course of this analysis. Instead, the algorithm was set to use ORFs annotated in the APP serovar 5 L20 genome as reference for APP mRNA partner searches. Putative GcvB sequences from both serovars have only a nucleotide mismatch at the 154 position, located outside its predicted seed regions (Figure 3.9), so unlikely to affect the prediction of its mRNA partners.

In this analysis, only predicted interactions between GcvB and its putative mRNA targets with the lowest hybridisation energy and scoring p-value less or equal to 0.05 were considered (standard value in TargetRNA2), resulting in 50 GcvB target mRNA candidates with energy scores between -18.62 and -8.35 kcal/mol. From this selection, all predicted targets displaying ORFs matched to non-annotated or hypothetical proteins were removed from the list leaving a selection of 30 targets (Appendix 7). This decision was made to increase the likelihood to relate these targets to annotated metabolic pathways in genomic databases and, consequently, to link the predicted sRNA to a regulatory pathway in APP.

3.3.2. Predicted interactome of putative targets

The remaining 30 predicted targets of APP8 GcvB were compared against STRING databases (Snel *et al.*, 2000; Jensen *et al*., 2009), using APP serovar 5 L20 genome for reference, to enquire whether their gene products share a function in similar metabolic pathways based on gene ontology (GO) terms. The output is illustrated in Figure 3.10 and enrichment values presented in the Appendix 6. Functional enrichment of 10 genes (*ilvD*,

leuC, *dapA*, *trpD*, *ilvC*, *dapE*, *hisG*, *dxs*, *ureA* and *gcp*) matched STRING annotation for biological processes of 11 GO-terms connected to amino-acid biosynthesis. This result aligned with validated experimental data from studies on GcvB in *Salmonella*, whose amino-acid transport biosynthesis regulon involves targeting of 45 mRNAs (Sharma *et al.*, 2011). From the 30 genes entered in the query, 15 of them share at least one functional association based on parameters such as co-expression, gene fusion experimental data (*ilvI*, *ilvE*, *ilvC*, *ilvD*, *thrC*, *leuC*, *serA*, *serB*, *serC*, *dxs*, *hisC*, *hisD*, *hisG*, *hisI* and *trpD*.

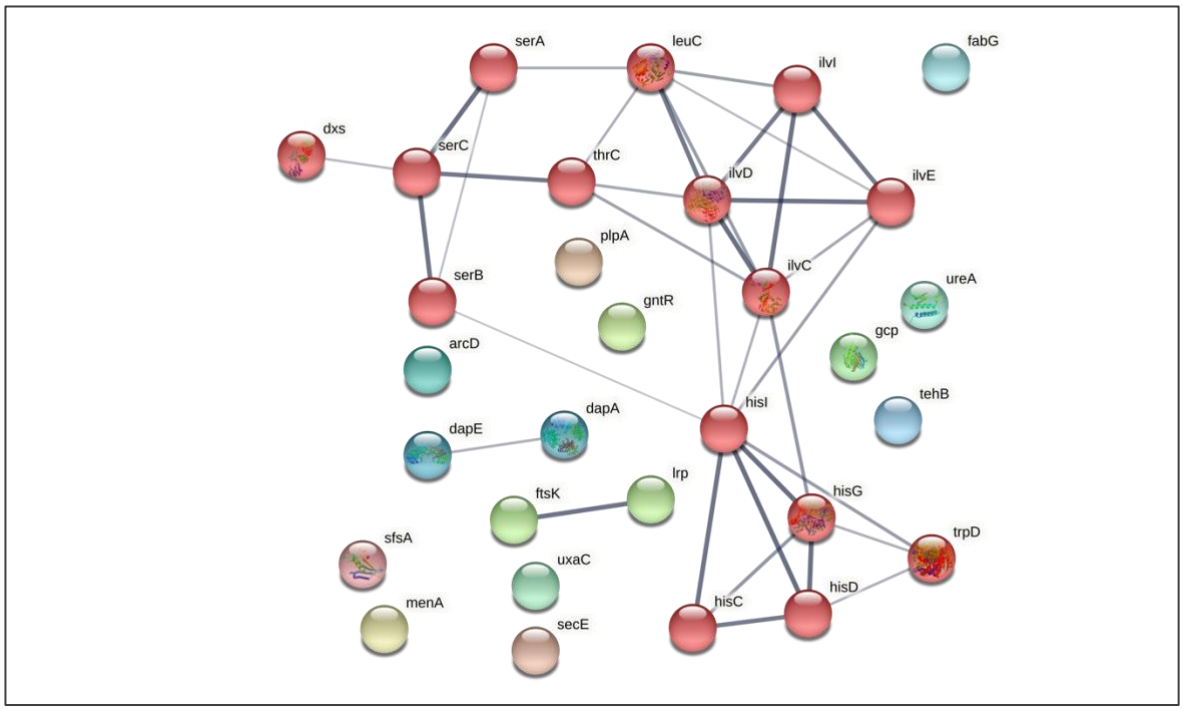


Figure 3.10. Protein-protein associations predicted by STRING for the predicted targets of APP8 GcvB. Grey lines represent protein-protein associations, and their thickness indicates the strength of data support. Associations are meant to be specific and meaningful, for example, proteins that jointly contribute to a shared function. This does not necessarily mean they are physically binding each other. Genes were analysed with a moderate confidence 0.400 and Markov clustering method (MCL) with inflation parameter 1.1. Interaction enrichments investigate whether proteins have more interactions among themselves than would be expected for a random set of proteins. Associated proteins share colour of the network and the red circles indicate proteins associated with amino acid biosynthesis. This set PPI enrichment p-value is < 1.0e-16. Such an enrichment indicates that the proteins are at least partially biologically connected, as a group.

When the same approach to predict and analyse mRNA targets was applied to the next putative *trans*-acting sRNA in the list (RNA01 in Table 3.2), STRING analysis of its selected mRNAs show few protein associations and a disperse functional enrichment (Figure 3.11). Rather than implying that RNA01 is not a plausible novel sRNA candidate because it lacks a major associated predicted metabolic pathway to regulate, this output may suggest that target prediction by the existent tools lack profiling information on novel sRNAs, as previously suggested in this work. For example, one of RNA01 targets is *ompP2*, the gene for the outer membrane protein P2, which forms pores that allow passive diffusion of small molecules across the outer membrane. This is an antigenic protein and has been identified as a potential virulence factor in *Haemophilus parasuis* (Zhang *et al.*, 2012). Targeting of *ompP2* suggests that RNA01 could be an interesting novel sRNA to be investigated using the RNA array to explore interaction with potential mRNA partners.

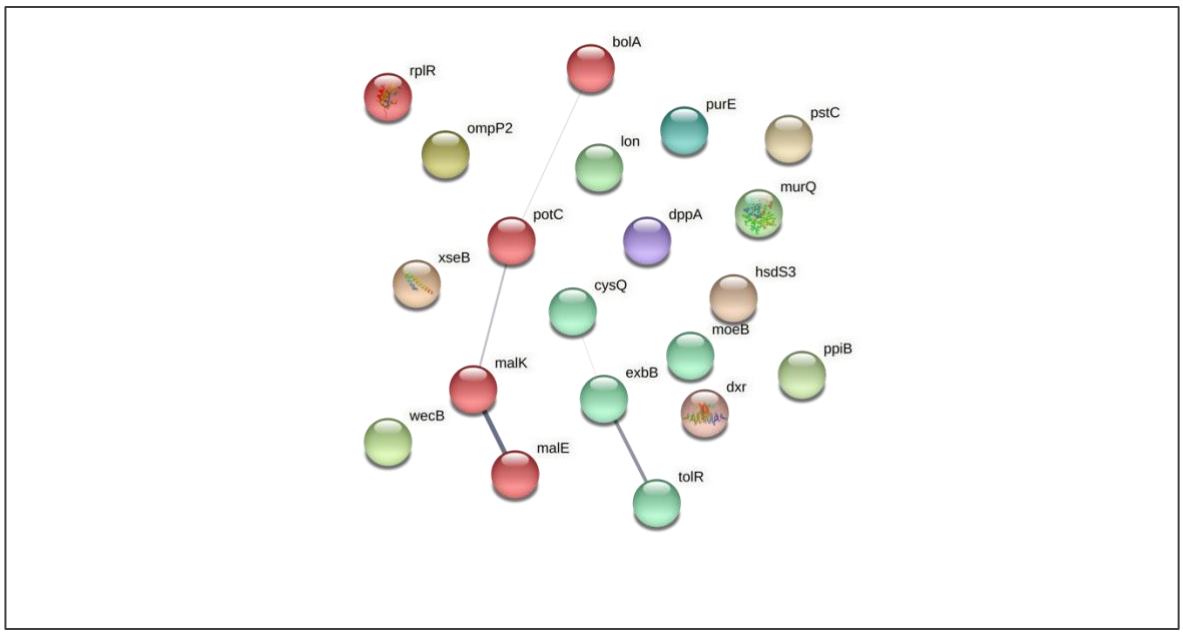


Figure 3.11. Protein-protein associations predicted by STRING of the predicted targets of RNA01. A quick query into the targets of the predicted novel sRNA RNA01 using the same parameters from Figure 3.10, shows a scattered protein-protein association distribution when compared to protein-protein associations for the predicted targets of APP8 GcvB. The enrichment p-value for the query is 0.989, suggesting that the predicted targets lack significant associations.

3.3.3. Selecting predicted targets using RNA-seq data

As previously mentioned, mRNA elements such as transcription start site (TSS) and 5' untranslated regions (5' UTR) may not be easily identified without the support of transcriptome insights provided by RNA-seq data. For this reason, a table (Appendix 7) was created containing basic information on the 30 predicted targets (gene name, description, free energy, p-value, interaction region and similarity to APP serovar 8 MIDG2331) and by adding information gathered from the RNA-seq TSS-enriched experiments to identify the TSS of the target transcripts, and ultimately the precise 5' UTR sequence. Data mining for this information was carried out in a similar manner used to select for novel *trans*-acting sRNAs in the beginning of this pipeline, by analysing expression cliffs of each transcript in JBrowse, and removing predicted targets displaying conflicting expression and lacking consistent TSS boundaries.

Other parameters included selecting for targets with identical nucleotide sequences in both APP genomes (L20 and MIDG2331), interaction with the predicted seed region of APP8 GcvB, and the location of the predicted binding regions in the putative targets. For example, sRNA-mRNA interaction regions in the mRNA are usually located around the ribosome binding site (RBS) or the start codon of the gene ORF. The RBS is located in the 5' UTR of the target transcript, thus predicted seed regions located outside the 5' UTR defined in the RNA-seq data, or more than 45 bases upstream the ORF data were removed from the selection. The 45 bases cut off was established to accommodate RNA array experimental design. Once these parameters were applied to the list of 30, the number of targets was then reduced to 9 mRNA target candidates of APP8 GcvB, including *ilvI, ilvE, thrC, serB, serC, tehB, ilvC, menA* and *hisG*. Their sequences, including 5' UTR boundary from RNA-seq, were extracted from GenBank for both APP serovar 5 L20 and 8 MIDG2331, to check for sequence similarity mismatches. Except for *menA*, all these sequences presented 100% similarity in both serovars, reducing the set to 8 mRNA targets. However, due to its role in pathogenicity in APP (Bossé & MacInnes, 2000), the predicted target ureA (urease gamma subunit) was added to the list, despite being part of an operon. The final list of plausible predicted targets of APP8 GcvB contains 9 targets (*ilvI, ilvE, thrC, serB, serC, tehB, ureA, ilvC* and *hisG*) to be taken forward in this study (Table 3.3).

Table 3.3.
List of predicted mRNA targets of the putative APP8 GcvB

List Rank *	Predicted Target	Description	Similiarity APP L20	Operon	RNA-seq TSS	Strand	5' UTR Location	Length (bp)	Seed Region mRNA	SD Location mRNA**	SD Location sRNA***
2	ilvC	ketol-acid reductoisomerase	100%	No	Consistent	-	LN908249.1:2148620..2148541	1482	TGTAAACACAACATCATAA	-62 to -44	72 to 90
3	ilvI	acetolactate synthase 3 catalytic subunit	100%	No	Consistent	+	LN908249.1:806311..806360	1725	AGCAAACACAATATTT	-41 to -26	75 to 90
5	thrC	threonine synthase	100%	No	Consistent	+	LN908249.1:1757818..1757859	1281	ACTTATACAACATCAA	+5 to +20	74 to 89
10	serC	phosphoserine aminotransferase	100%	No	Consistent	-	LN908249.1:777644..777609	1089	GGCAAACACAATGACA	-10 to +6	74 to 90
14	ureA	urease subunit gamma	100%	Yes	Consistent	-	LN908249.1:1892798..1892766	303	ACAAACACAAACAA	-28 to -15	77 to 89
17	serB	phosphoserine phosphatase	100%	No	Consistent	+	LN908249.1:1373092..1373140	861	TTATGCCAAACACGATT	-2 to +15	78 to 93
21	hisG	ATP phosphoribosyltransferase	100%	No	Consistent	+	LN908249.1:2308592..2308636	900	TTAAACACAACATGGG	-25 to -10	74 to 89
26	ilvE	branched-chain amino acid aminotransferase	100%	No	Consistent	-	LN908249.1:84030..83991	1028	GGACACAACAACATGGCTTT	-12 to +8	70 to 88
29	tehB	tellurite resistance protein TehB	100%	No	Consistent	+	LN908249.1:1584536..1584566	873	GATTAATTATGCAAAACT	-8 to +10	81 to 98

* Likelyhood of regulatory target prediction ranked by P-values in RNATarget2
** Location of putative binding in the mRNA predicted in RNATarget2
*** Location of predicted interaction region in the sRNA predicted in RNATarget2

The list rank is ordered by the lowest hybridisation energy and p-values (Appendix 7) to show the most likely mRNA partners of APP8 GcvB. Description of targets suggests that most are directly involved in amino-acid biosynthesis. The parameters similarity, operon, RNA-seq TSS, interaction region distance on the mRNA and seed region position in the sRNA were used to select the most consistent and structurally plausible candidates from the initial TargetRNA2 output of 50 candidates. Some targets have their interaction region nearby or overlapping the start codon of the gene coding region +1 residue (*thrC, serC, serB, ilvE* and *tehB*), while others (*ilvC, ilvI, ureA* and *hisG*) sit on the 5'UTR within -62 nucleotide position. These regions are C/A-rich sequences. All targets are predicted to preferably interact with APP8 GcvB within the boundaries of its predicted seed region R1 (78-90, sequence 5' – UGUUGUGUUUGCA – 3', CV1 in alignment Figure 3.7).

3.3.4. The predicted mRNA targets of the putative GcvB are involved in pathogenicity

From the selection of candidates, most are associated with amino-acid biosynthesis pathways (*ilvE, ilvI, ilvC, thrC, serB, serC*, and *hisG*). Genes *ilvC, ilvE* and *ilvI* are involved in in the metabolic pathway that synthesizes L-isoleucine and L-valine, while *serB* and *serC* are involved in a sub-pathway that synthesizes L-serine. Gene *thrC* synthesizes L-threonine and *hisG* L-histidine. Genes involved in amino-acid biosynthesis contribute to bacterial pathogenicity by supporting survival and infection persistence. Bacteria unable to synthesize branched-chain amino acids may be attenuated, as suggested in a study with deletion-disruption mutants of the *ilvI* in APP by Subashchandrabose *et al.* (2009).

The predicted target *ureA*, a gene coding for the urease subunit gamma, is part of a sub-pathway that synthesizes CO_2 and NH_3 from urea, which is part of the urea degradation pathway. Urease is a known virulence factor characterised in APP strains, whose activity is involved in avoiding the host's defence mechanisms by impairing macrophage function (Bossé *et al.*, 1997; 2000; 2001). The putative GcvB in APP may regulate expression of a mRNA from a gene (*tehB*) that codes for a tellurite resistance protein (S-adenosyl methyltransferase), involved in resistance to oxidative stress in APP (Subashchandrabose *et al.*, 2013), therefore supporting bacterial pathogenicity.

3.3.5. Putative GcvB sRNA may bind predicted mRNA targets on A/C-rich sequences

Salmonella GcvB binds to extended C/A-rich regions in its mRNA targets (Sharma *et al.*, 2007). Alignment of the predicted preferable binding regions of the APP8 GcvB mRNA targets predicted by TargetRNA2 (Figure 3.12), shows that the varied A/C-rich repeats in all the mRNA targets may form a consensus order AAACACAA, despite random mismatches in each target. Imperfect alignment goes in line with imperfect pairing of sRNA-mRNA interactions, a mechanism that may have been conserved/evolved to accommodate for genomic alterations such as point mutations, gene duplication and horizontal gene transfer among others, as one sRNA can regulate a plethora of targets (reviewed in Updegrove *et al.*, 2015).

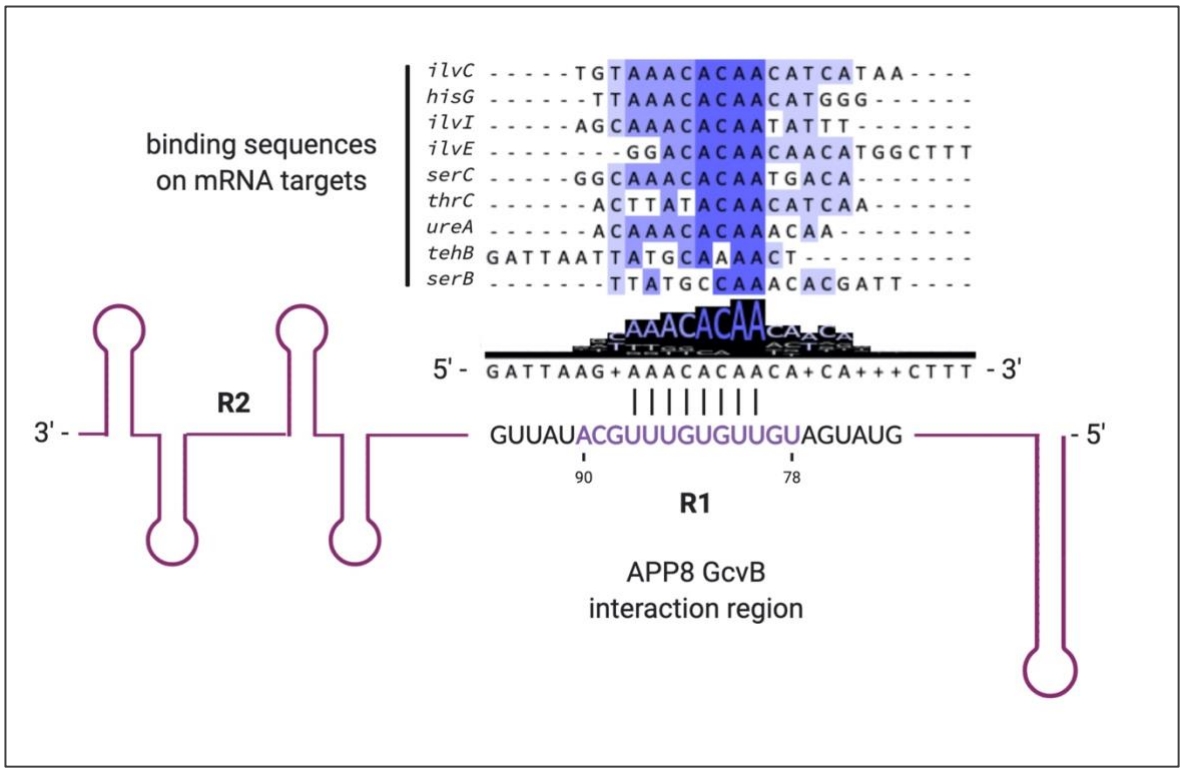
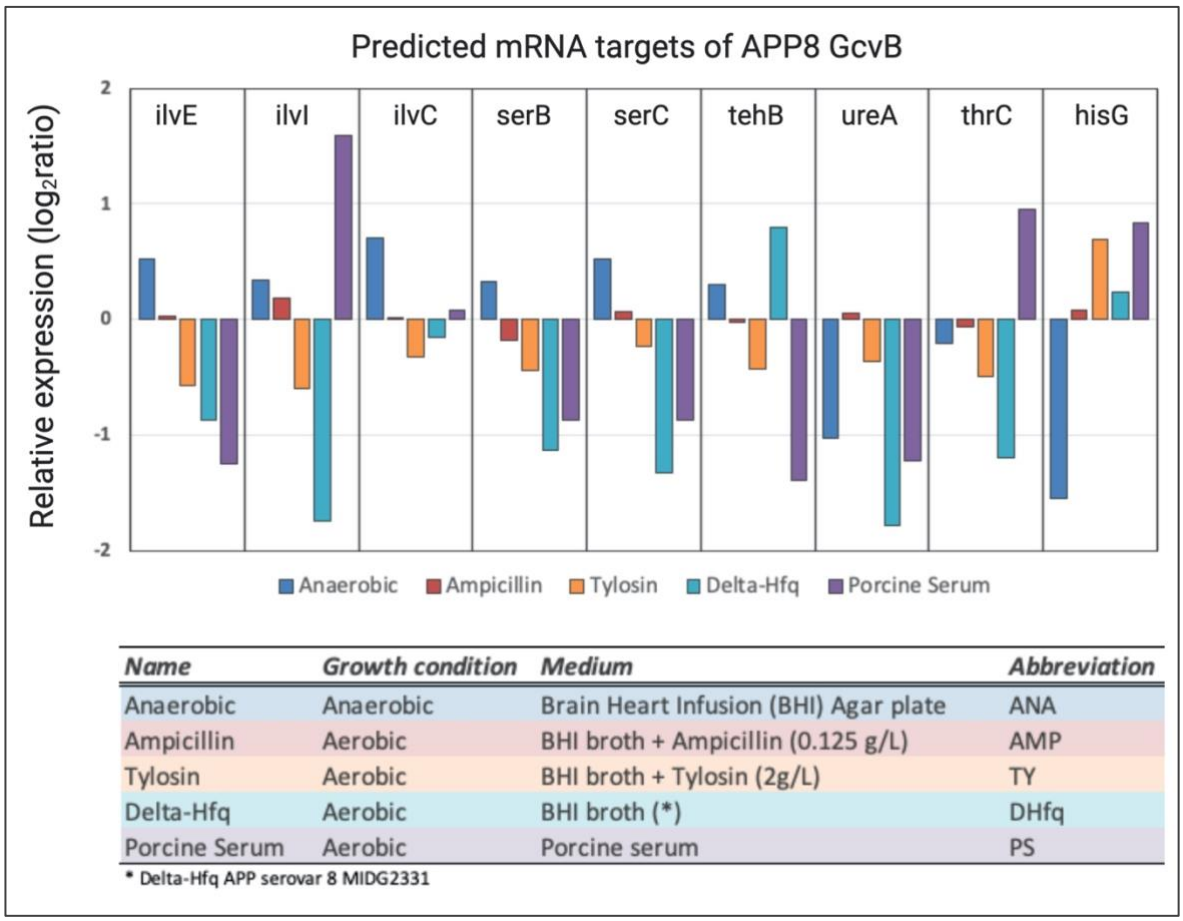


Figure 3.12. Alignment of the predicted APP8 GcvB mRNA target binding regions. The predicted binding regions of the selected mRNA targets were aligned using ClustalOmega and visualised by JalView. Percentage identity was coloured in purple gradient and consensus sequence histogram displayed at the bottom of the alignment. Schematics of APP8 GcvB predicted structure with 5 stem-loops and the sequence of the first conserved region (R1) was designed in purple in the opposite direction (3' to 5') of the mRNAs (5' to 3') to illustrate sRNA-mRNA hybridisation, based on an imperfect complementary base-paring interaction. The highest alignment consensus sequence (AAACACAA) complementarily matches a region in the centre of APP8 GcvB R1 (location 79 to 90 in the sRNA) without sequence mismatches. See Appendix 7 for intermolecular interaction energy in kcal/mol.

3.3.6. Expression analysis of the predicted mRNA targets

Expression analysis of the selected predicted mRNA targets of APP8 GcvB was carried out following the same rationale used for the analysis of the 9 selected P-sRNAs (section 3.2.4). As previously mentioned, transcript measurements in this study do not represent their differential expression, but rather the transcript yield of MIDG2331 in various growth conditions in relation to the aerobically grown MIDG2331 wild type (Figure 3.13).



Name	*Growth condition*	*Medium*	*Abbreviation*
Anaerobic	Anaerobic	Brain Heart Infusion (BHI) Agar plate	ANA
Ampicillin	Aerobic	BHI broth + Ampicillin (0.125 g/L)	AMP
Tylosin	Aerobic	BHI broth + Tylosin (2g/L)	TY
Delta-Hfq	Aerobic	BHI broth (*)	DHfq
Porcine Serum	Aerobic	Porcine serum	PS

* Delta-Hfq APP serovar 8 MIDG2331

Figure 3.13. Relative transcript expression of the predicted mRNA targets of APP8 GcvB. Transcript yield relative to WT in broth or Agar plate media was calculated and normalised ($Ratio_2log$) for all 9 predicted mRNAs. Expression levels of MIDG2331 wild type (WT) in **anaerobic** (ANA) BHI Agar plate was compared against WT in **aerobic** growth condition in Agar plate (blue bars). MIDG2331 WT with added **AMP** (red bars) and **TY** (orange bars) experiments were carried out in BHI broth in aerobic growth conditions and both compared to MIDG2331 WT grown in aerobic BHI broth. The yield of MIDG2331 Hfq knockout (**DHfq**) in BHI broth (green bars) was compared against MIDG2331 WT grown in BHI broth. Finally, transcript ratio of MIDG2331 WT grown in porcine serum (PS) was compared against MIDG2331 WT grown in aerobic BHI broth (purple bars).

The graph shows that targets *ilvE, ilvI* and *ilvC*, involved in the synthesis of L-isoleucine and L-valine, have similar transcript yield patterns when grown in anaerobic (~> +0.5-fold), AMP (~> +0.1-fold), TY (~< -0.3-fold) and DHfq (~< -0.1-fold) conditions in contrast to wild type aerobic. This group of genes only have different positive and negative yield trends in PS, with *ilvI* presenting the highest yield for this condition (~ +1.5-fold) among all other targets, while *ilvC* shows little change (~ +0.1-fold) in PS. Targets *serB* and *serC*, involved in the synthesis of L-serine, also have similar yield trends, except for AMP condition, with *serB* demonstrating a negative yield (~ -0.2-fold) and *serB* slightly positive (~ +0.1-fold). The remaining targets belonging to different sub pathways show a very varied trend instead, with *ureA* displaying very low transcript yields in most conditions, except for a slightly positive fold in AMP (~ +0.1-fold). Target *thrC* implies a similar negative fold pattern in most conditions, except for PS, with nearly the double of total transcripts of this gene in wild type. Finally, *hisG* shows positive transcript fold in most conditions, except for anaerobic (~ -1.5-fold) when compared to wild type grown in aerobic conditions.

An overall glimpse at the graph shows that, generally, most targets had a decreased transcript fold in most conditions. However, although this data provides a snapshot of the transcript yields of the studied sRNAs and targets at the time of their extraction, it offers no information on sRNA-mRNA interactions, or whether these mRNAs are in fact being targeted and regulated by APP8 GcvB under these conditions. This reinforces the need for the strategy created in this project, to couple increased prediction accuracy with *in vitro* testing of these sRNA-mRNA pairs using the RNA array technology.

3.3.7. Expression map of APP serovar 8

While in this work only a selected group of genes were investigated, the software Circos allowed a global analysis of gene expression in MIDG2331 by simultaneously comparing transcript yields of all its genes in different bacterial growth conditions in a circular graph. Novelty is one of the main reasons to use this visualisation tool, which identifies generalised changes in expression patterns. By closely inspecting the Circos graph

in Figure 3.14, is possible to identify localised transcript yield discrepancies in regions of the whole MIDG2331 genome for a variety of conditions at once.

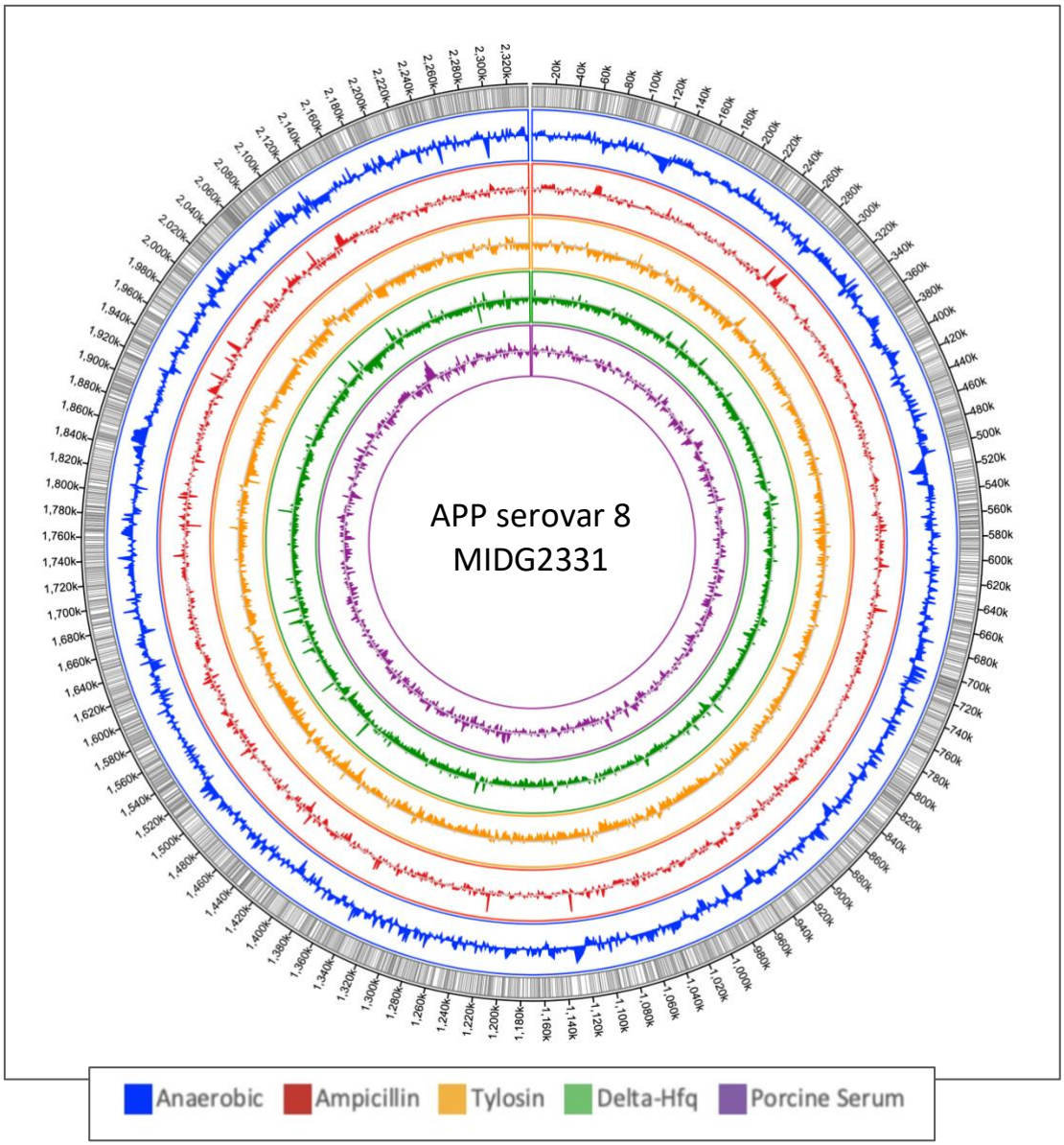


Figure 3.14. Circular graph of MIDG2331 transcript yield. Transcript yield related to expression levels of MIDG2331 wild type (WT) in **anaerobic** (ANA) BHI Agar plate was compared against WT in **aerobic** growth condition in Agar plate (blue). MIDG2331 WT with added **AMP** (red) and **TY** (orange) experiments were carried out in BHI broth in aerobic growth conditions and both compared to MIDG2331 WT grown in aerobic BHI broth. The yield of MIDG2331 Hfq knockout (**DHfq**) in BHI broth (green) was compared against MIDG2331 WT grown in BHI broth. Finally, transcript ratio of MIDG2331 WT grown in **porcine serum** (PS) was compared against MIDG2331 WT grown in aerobic BHI broth (purple).

3.4. Summary and Discussion

The *in silico* pipeline used in this chapter has identified novel *trans*-acting sRNA candidates in APP with potential links to virulence. Comparative studies have confirmed that the predicted *trans*-acting sRNA RNA05/ARRC01 has a high identity to GcvB from *E. coli* and *Salmonella*, which are both well characterised (Urbanowski *et al.*, 2000; Sharma *et al.*, 2007; Miyakoshi, 2015). GcvB is known to repress the mRNAs of proteins involved in amino acid uptake and synthesis; processes which are involved in bacterial virulence. The pipeline used here identified mRNA targets for RNA05/ARRC01 which are connected to amino acid biosynthesis pathways.

Whilst GcvB has been identified in *Pasteurella multocida*, a closely related pathogenic member of the *Pasteurellaceae* family (Gulliver *et al.*, 2018), gene ontology studies for APP serovar 5 and *Salmonella* have also identified several putative target mRNAs (including *ilvC*, *ilvI* and *ilvE*) which are involved in amino acid biosynthesis and associated with virulence (Chiers *et al.*, 2010; Sharma *et al.*, 2011; Rossi *et al.*, 2016). Therefore, the observation that the predicted mRNA targets for RNA05/ARRC01 include *ilvC, ilvI, ilvE*, and others linked to amino acid biosynthesis pathways, which impact virulence, gives confidence in suggesting the *trans*-acting sRNA RNA05/ARRC01 is GcvB in APP8. Importantly, the other remaining *trans*-acting sRNAs in the UoP_ICL_UoV_2020 list showed no homology to other species, making them completely novel sRNAs across species.

RNA-seq data from collaborators was used to allow comparison of transcript expression yields across the different growth conditions. However, it was not possible to directly convert these measurements to differential expression analysis. Statistical methods cannot remove noise from experiments which were not designed for this purpose and, although these experiments confirm that transcripts for the putative sRNAs and mRNA targets are present in APP8 at a determined condition, whether those are in fact sRNAs interacting with mRNA targets is yet to be confirmed. This set of data provided only a snapshot of the APP MIDG2331 transcriptome, which is not a definitive rate of gene expression throughout host infection, but a picture of which sRNAs are present at the stage of RNA extraction.

Transcript yield may provide insights into mRNA levels linked to a given metabolic pathway during pathogenicity, as expression of different genes are turned on and off during the stages of bacterial infection, affecting the number of transcripts of a gene. For example, the genes analysed by STRING in section 3.3.2 Figure 3.10 are associated with amino acid biosynthesis and looking at their expression yields in different growth conditions shines a light on their presence in relation to GcvB.

RNA-seq data on GcvB suggests a higher transcript yield when bacteria grow in nutrient-rich media such as PS. In this context, a possible suggestion is that GcvB may be regulating its targets to optimize the energetically expensive biosynthesis of amino acids (Sharma *et al.*, 2007) by reducing their transcript yield. Apart from *ilvI*, *thrC* and *hisG* transcripts, yield is reduced for all other mRNA targets when APP is grown in this media in relation to BHI broth wild type.

Collectively, the work in this chapter has output a putative *trans*-acting sRNA, proposed to be APP8 GcvB, as well as its predicted mRNA targets. Whilst the RNA-seq data is suggestive of biological relevance, it cannot provide interaction information. By contrast, computational analysis as part of the pipeline has predicted the sRNA-mRNA binding regions, and relative binding energies. In the next chapter, *in vitro* testing of these sRNA-mRNA pairs will be undertaken using in the RNA array technology to explore the interactions experimentally (Chapter 4).

Chapter 4

4. *In vitro* studies: Analysis of virulence-associated sRNA-mRNA interactions from APP

4.1. Introduction

In Chapter 3, novel *trans*-acting sRNAs that are linked to virulence in APP serovar 8 MIDG2331 were identified using RNA-seq data, and bioinformatic methods were used to predict their putative mRNA targets. This chapter establishes an experimental pipeline that could be used to test the binding interaction between these predicted sRNA-mRNA pairs *in vitro*. Among the identified APP sRNAs, APP8 GcvB (predicted sRNA ARRC01/RNA05 in Chapter 3, referred to as GcvB in this chapter) and its selected putative mRNA targets (*ilvI, ilvE, ilvC, serB, serC, tehB, ureA* and *hisG*) stand out as viable first candidates to be investigated. This is because GcvB and its mRNA targets have been characterised in other organisms (Urbanowski *et al.*, 2000; Sharma *et al.*, 2007; Miyakoshi, 2015) and its function in APP is supported by experimental data from colleagues at Imperial College London and Universidade de Viçosa in Brazil (Rossi *et al.*, 2016).

As previously covered in Chapter 1 (section 1.2.1), bacterial *trans*-acting sRNAs regulate gene expression by base-pairing to specific sequences of their target mRNAs. This base-pairing interaction typically regulates translation of the mRNA by modulating the accessibility of the mRNA ribosomal binding site (RBS) to ribosomes. The sRNA-mRNA base-pairing interaction consists of partial sequence complementarity over a short stretch of sequence, and *in vivo*, the interaction is often facilitated by Hfq. The ternary sRNA-mRNA-Hfq complex, and its mechanism-of-action, have been extensively studied elsewhere for *E. coli* and *Salmonella* (Vogel & Luisi, 2011; Henderson *et al*., 2013; Westermann *et al*., 2019) and will not be investigated here. Instead, this study focusses on the direct interaction between an sRNA and its mRNA targets, which is frequently observed *in vitro*, even in the absence of Hfq (Henderson *et al.*; Vincent *et al.*, 2013).

The aim of this chapter was to investigate whether GcvB can interact with its putative mRNA targets (*ilvI, ilvE, ilvC, serB, serC, tehB, ureA* and *hisG*) *in vitro* using the novel RNA array technology developed at the University of Portsmouth by the Callaghan group. The RNA array technology is an adaptable technique that can be used to investigate a variety of RNA based interactions (Phillips *et al*., 2018; Norouzi *et al*., 2019; Henderson *et al*., 2019). As discussed in Chapter 1 (section 1.5) and Chapter 2 (section 2.2.1), the first step is to generate an RNA array through a method that involves *in vitro* transcription (IVT) of a DNA IVT template array and *in situ* surface capture of the RNA on a facing surface. The RNA array can then be used to investigate RNA-RNA binding (Phillips *et al*., 2018, Norouzi *et al*., 2019) or RNA-small molecule binding (Henderson *et al*., 2019).

This chapter first describes the design of the DNA IVT templates for each of the APP GcvB putative mRNA targets. It then details the production of the DNA IVT template array on slide and the visualisation and analysis of the levels of DNA on the DNA IVT template array. Next, the generation of the RNA array of mRNA targets by IVT and *in situ* RNA capture is presented, together with the visualisation and analysis of the levels of RNA on the mRNA target array. Quantification of the DNA and RNA levels on the respective arrays then allowed for an investigation into the relative transcription efficiency of each of the mRNA targets. Finally, two strategies were devised to screen the mRNA targets for GcvB binding. The first involved probing the mRNA target array with GcvB post-array production and the second involved co-transcribing GcvB with the putative mRNA targets. The workflow for this *in vitro* RNA array-based sRNA-mRNA interaction validation pipeline is illustrated in Figure 4.1.

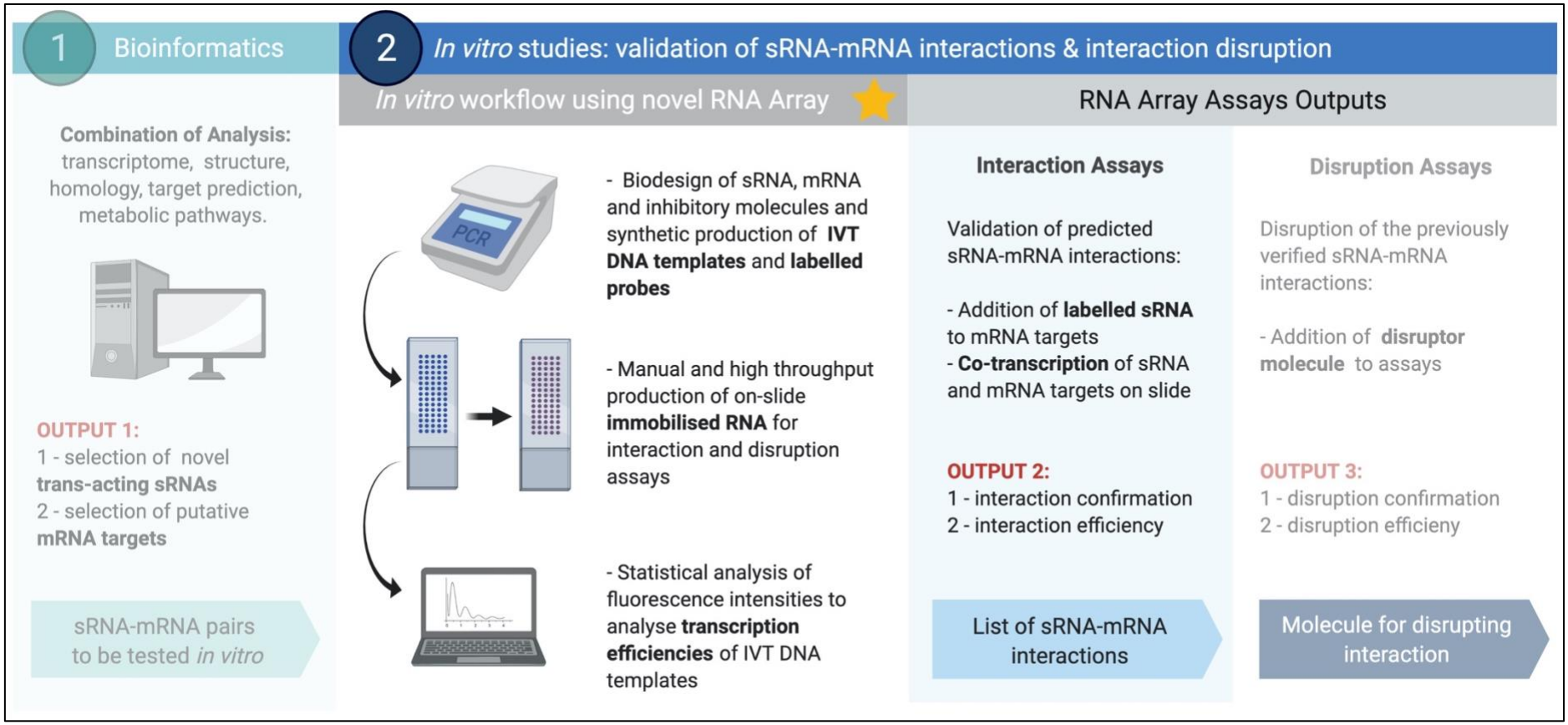


Figure 4.1. Developing an *in vitro* RNA array-based sRNA-mRNA interaction validation pipeline. The first stage of the *in vitro* RNA array-based sRNA-mRNA interaction validation pipeline is the biodesign and synthesis of a DNA IVT template array. The DNA IVT template array is used to generate an RNA array of mRNA targets by IVT and *in situ* RNA capture. A fluorescent label incorporated into the immobilised DNA IVT templates, and a fluorescently labelled oligonucleotide probe that binds specifically to the immobilised mRNA targets, allows for detection and quantification of the immobilised DNA IVT template and mRNA target levels. Analysis of the DNA IVT template and mRNA target levels is used to calculate transcription efficiencies. Addition of the sRNA to the mRNA target array allows for validation of sRNA-mRNA interactions.

4.2. Design of the DNA IVT templates and generation of DNA IVT template arrays

4.2.1. Introduction

The first step in the *in vitro* RNA array-based sRNA-mRNA interaction validation pipeline was the design and synthesis of the DNA IVT templates, and the production of the DNA IVT template array. This section describes the key elements of the DNA IVT template design, both manual and automated methods for producing the DNA IVT template array and the strategy for visualisation and quantification of the DNA IVT template levels on the DNA IVT template array.

4.2.2. DNA IVT template design and synthesis

DNA IVT templates were designed for each of the nine GcvB putative mRNA targets. The key features and sequence elements of the DNA IVT templates are described in Figure 4.2 and stablished in the RNA array standard protocol (Phillips *et al*., 2018, Henderson *et al*., 2019, Vincent *et al*., 2021 *in press*). Each DNA IVT template consists of a series of functional sequence features that are required for producing the DNA IVT template array, producing the RNA array or visualising the arrays. These include: 1 – A biotinylated linker at the 5' end to facilitate DNA IVT template immobilisation onto a streptavidin (SA)-coated slide; 2 – A T7 bacteriophage RNA polymerase (RNAP) promoter for IVT; 3 – A three guanine (GGG) nucleotide repeat to optimise transcription by the T7 RNAP; 4 – Sequence encoding the RNA of interest (ROI); 5 – Sequence encoding an unstructured linker (SA-linker) to physically separate the mRNA target and the downstream streptavidin-binding RNA aptamer (SA aptamer) and to serve as the recognition region for a fluorescently labelled antisense single-stranded DNA (ssDNA) probe used for visualisation and quantification of the mRNA target levels on the RNA array; 6 – Sequence encoding the SA aptamer to facilitate *in situ* capture of the transcribed RNA by an SA-coated capture slide. 7 – A

fluorophore at the 3′ end to allow visualisation and quantification of the DNA IVT template levels on the DNA IVT template array.

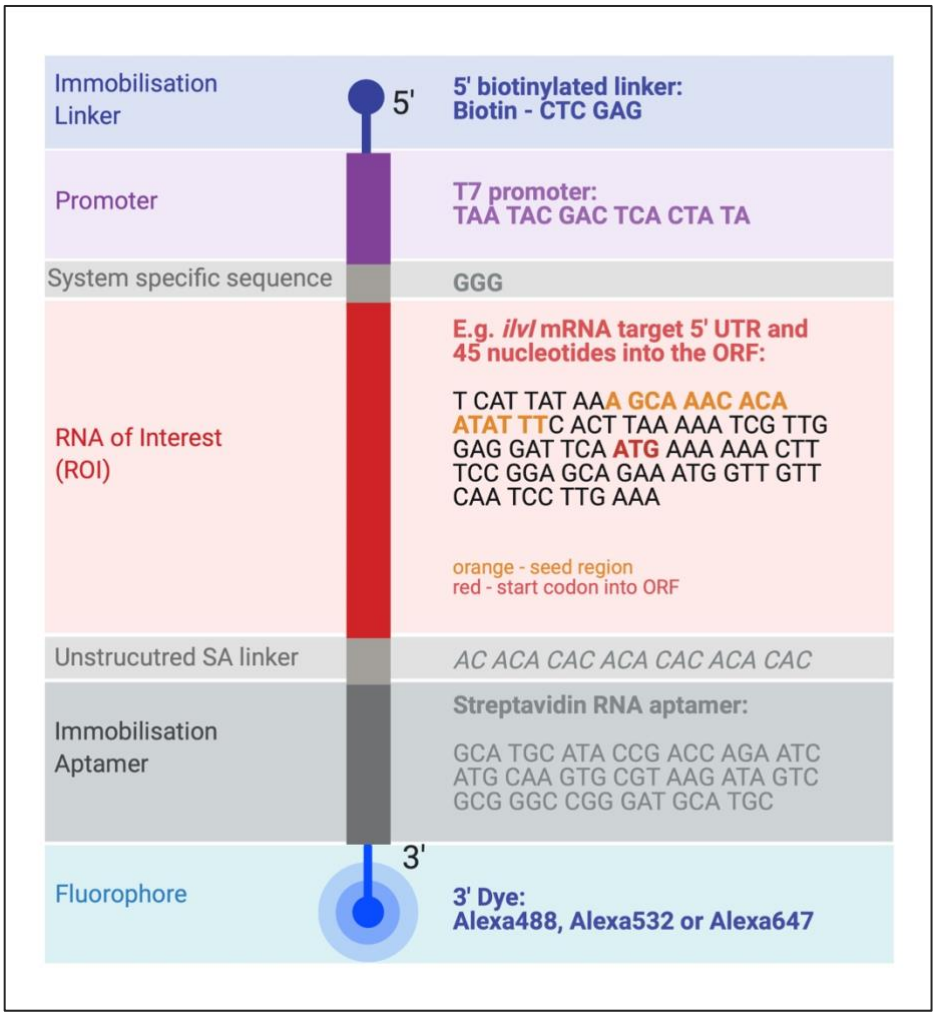


Figure 4.2. Design of the DNA *in vitro* transcription templates. A schematic of the DNA IVT template. The DNA sequence for each component is indicated. An example RNA of interest (ROI) sequence is given for the *ilvI* mRNA target. This consists of the predicted GcvB-interacting region within the 5′ UTR (orange – seed region) and 45 nucleotides of the ORF. The final 201-nucleotide sequence for the *ilvI* DNA IVT template is: 5′-biotin-ctcgagTAATACGACTCACTATA*GGG*TCATTATAA**AGCAAACACAATATTT**CACTTAAAAATCGTTGGAGGATTCA**ATG**AAAAAACTTT CCGGAGCAGAAATGGTTGTTCAATCCTTGAAA*ACACACACACACACACACAC***GCATGCATACCGACCAGAATCATGCAAGTGCGTA AGATAGTCGCGGGCCGGGATGCATGC**-Alexa(488, 532 or 647)-3′. Sequences for the DNA IVT templates for the other mRNA targets are given in Appendix 9.

The RNAs of interest (ROIs) used in this chapter are not the full-length mRNA targets. Each ROI consists of the 5′ untranslated region (UTR) of the GcvB putative mRNA target, defined by the TSS-enriched RNA-seq data from Chapter 3, and 45 nucleotides (15 codons) of the mRNA target's open reading frame (ORF). This region of the mRNA target includes the region predicted to base-pair with GcvB and also the functional sequence

features, such as A/C-rich regions, which are known to be required for target recognition. The sequences of the mRNA target ROIs are presented in Appendix 9. The DNA IVT templates are named based on the mRNA target ROI. For example, the DNA IVT template shown in Figure 4.2 is the mRNA target *ilvI* DNA IVT template and will be referred to simply as *ilvI*. When referred to the mRNA target ROI generically, it will simply be referred to as the mRNA or mRNA target.

In order to bind GcvB, the mRNA must fold correctly and, in order to facilitate capture by the SA-coated capture surface, the SA aptamer must fold correctly. Therefore, RNAFold (Gruber *et al.*, 2008) was used to predict the secondary structure for the expected RNA product for each of the DNA IVT templates to confirm that conjugation of the mRNA to the SA aptamer does not affect the structure of either RNA. For each RNA product the mRNA and the SA aptamer were predicted to fold as expected (Appendix 10).

Following DNA IVT template design, gene synthesis primers were designed for each of the IVT DNA templates and obtained from Invitrogen (Thermo Fisher Scientific). The IVT DNA templates were synthetised *de novo* in-house by TBIO-PCR (Chapter 2, section 2.2.3.2). The 5′ biotin and the 3′ fluorophore were incorporated using a 5′-biotinylated primer and a 5′-fluorescently labelled reverse primer. The size of each DNA IVT template was verified by gel electrophoresis (Appendix 12). Each DNA template produced an IVT product of the expected length in solution (Appendix 13).

4.2.3. Generating the DNA template array: spotting methods

Having designed and synthesised the 5′-biotinylated and 3′-fluorescently labelled DNA IVT templates, the next step was to produce a DNA IVT template array. This was achieved by spotting the DNA IVT templates onto an SA-coated slide so that the DNA IVT templates are immobilised via their 5′ biotin. Two spotting methods were employed, manual spotting, using a pipette, and automated spotting, using an arrayer robot. Schematics of manual and automated spotting of the IVT DNA templates are illustrated in Figure 4.3.

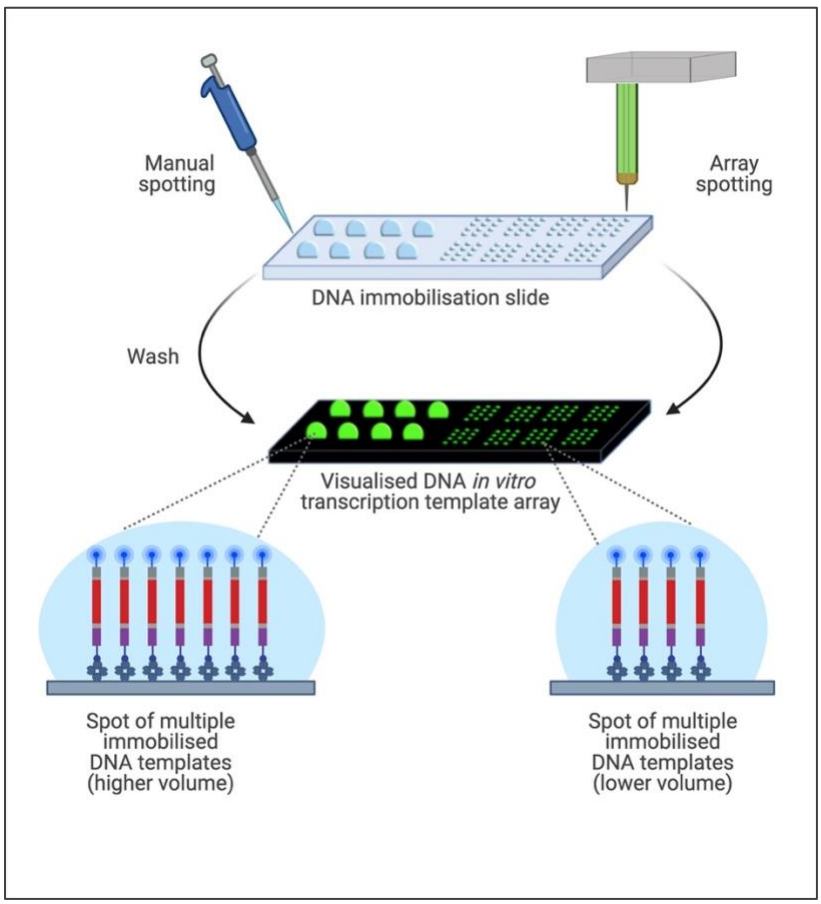


Figure 4.3. Spotting and visualisation of the DNA IVT template array. The 5′-biotinylated and 3′-fluorescently labelled DNA IVT templates are spotted onto an SA-coated slide by manual spotting with a pipette or by automated spotting using an arrayer robot. The top slide (blue) represents the spotting pattern and volume achieved for each spotting method. Manual spotting requires a higher volume of DNA IVT template than automated spotting. For example, 1 µl of DNA IVT template can be used to generate 5 spots by manual spotting or 130 spots by automated spotting. The middle slide (black) illustrates visualisation of the spots using a slide scanner following a wash step to remove any excess unbound molecules. Both methods result in discrete spots containing multiple molecules of the immobilised DNA IVT template, with the larger spots containing a higher level of immobilised DNA IVT template compared to the smaller spots, for the same solution spotting concentration.

Manual spotting allows for the generation of low- to medium-density arrays. It is possible to pipette as little as 0.2 µl per spot, resulting in 24 spots per slide (Phillips *et al.*, 2018). Therefore, manual spotting was typically used for preliminary experiments e.g., to check the functionality of a newly synthesised DNA IVT template or to test an experimental strategy when a yes/no output was all that was required. However, it is difficult to produce uniform spots by manual spotting.

On the other hand, automated spotting generates high-density arrays. For example, the Qarray2 arrayer robot (Genetix), fitted with a 200 µm pinhead and spotting at a spot

separation of 1250 µm results in a spot density of approximately 400 spots per slide (Vincent *et al.*, 2021 *in press*; Appendix 15). Furthermore, the spots produced by automated spotting are more homogenous with regard to volume, size and shape. Automated spotting was used for all high throughput experiments in this thesis.

4.2.4. Visualisation and quantification of the DNA IVT template levels on the DNA IVT template array

It is useful to visualise and quantify the levels of DNA IVT template immobilised on the slide, both as a quality control checkpoint to ensure that the DNA IVT template array has been generated successfully, but also for downstream analysis such as calculating the transcription efficiency of a DNA IVT template. To enable visualisation and quantification, the DNA IVT templates are designed and synthesised with a fluorophore incorporated at their 3′ end. Fluorophore fluorescence can then be detected using a microarray scanner which reports the emitted fluorescence from the fluorophore as fluorescence intensity (FI), in arbitrary units (a.u.). Although FI does not provide information about the absolute amount of DNA on the array, it is directly proportional to the amount DNA present and can be used to determine the relative amounts of DNA present between different array spots.

4.2.4.1. Fluorescence intensity (FI) measurements

In order to accurately measure the FI for a DNA IVT array spot, it is important to carefully define the boundaries of an array spot. The integrated scanner software can automatically recognise spots on a DNA IVT template array (Figure 4.4). The software then measures the FI for each pixel inside each spot boundary and also for each pixel outside the boundaries of the spots. FI for pixels outside the boundaries of the spots is considered to be background and can be subtracted from the spot FI. There are various possible sources of background fluorescence including auto-fluorescence of the slide or non-specific binding of fluorescently labelled molecules/probes.

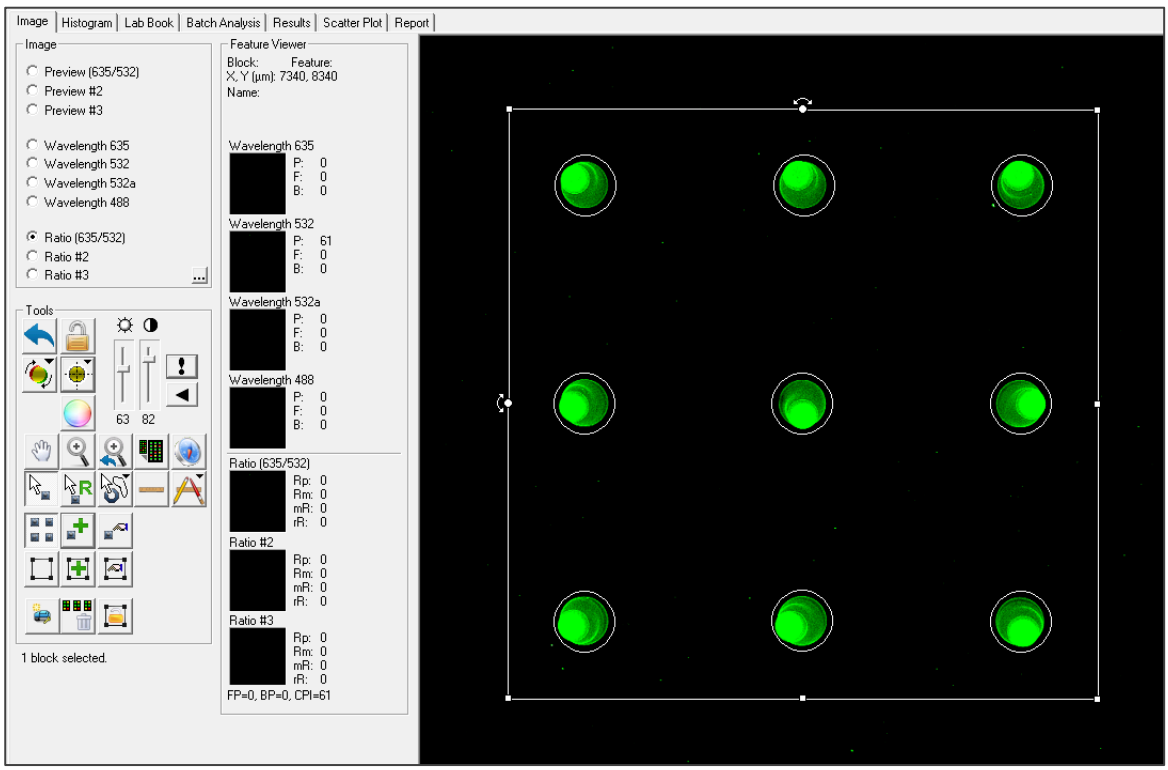


Figure 4.4. Spot visualisation using the integrated scanner software (GenePix). A screenshot of an array of 3′-AlexaFluor532 labelled DNA IVT templates visualised using an excitation wavelength of 532 nm and a Standard Green emission filter in the integrated scanner software. The panel at the left-hand side of the software interface contains excitation wavelength options, Tools for feature editing to position the spot boundaries, and the Feature Viewer that displays the FI for the different excitation wavelengths by pixel, the total FI for the spot and the background FI. Statistical analysis can be accessed from the tabs at the top of the panel, including the Results tab, where a table containing the median, mean and total FI for each feature can be found. The spots are shown in the panel on the right-hand side of the interface. White lines mark the spot boundary and the array field boundary.

4.2.4.2. Median, mean and total Fluorescence Intensity (FI)

The scanner software offers several options for statistical analysis of the FI. These include calculation of the total FI, median FI and mean FI. Total FI is the sum of the FI for every pixel inside the spot boundary, the median FI considers the distribution of FI for pixels inside the spot boundary and the mean FI is the total FI for the spot divided by the number of pixels inside the spot. The software also calculates the mean background FI and there is the option to automatically subtract the mean background FI from the mean FI. For the data reported in this thesis, the mean FI with the mean background FI automatically subtracted in the scanner software were used to compare spots.

4.2.4.3. Defining array spot boundaries with regular and irregular spot shapes

As mentioned above, in order to accurately measure the FI for a DNA IVT template array spot, it is important to carefully define the boundary of the spot. The scanner software allows the spot to be defined with a regular feature shape or an irregular feature shape (Figure 4.5). With a regular circle feature shape (Figure 4.5, 1A), in order to ensure that all of the spot FI is included, background FI is inadvertently included in the mean FI calculation. In contrast, with the irregular feature shape (Figure 4.5, 1B) only spot FI is included in the mean FI calculation.

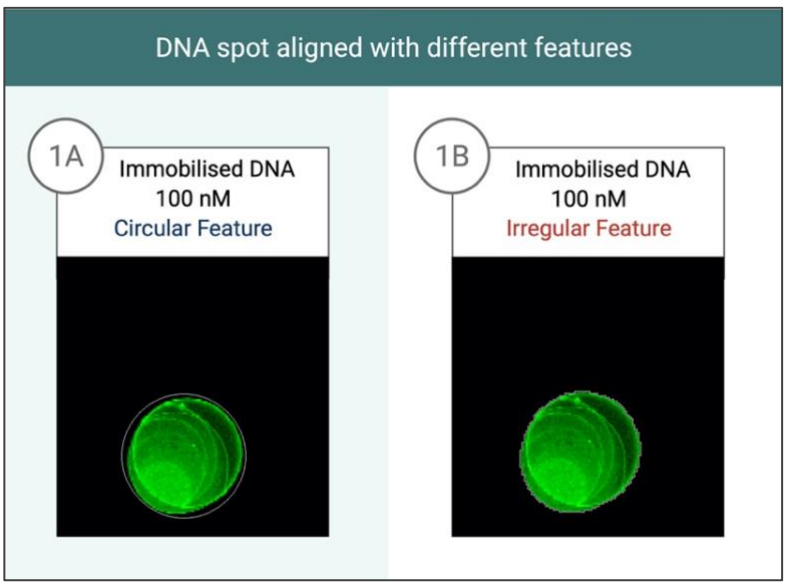
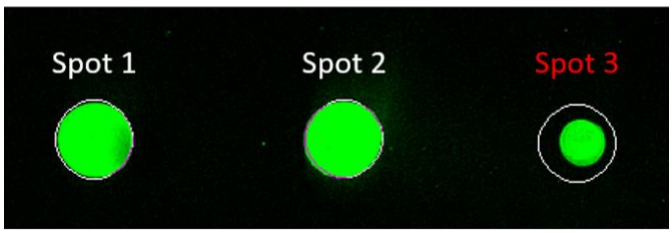


Figure 4.5. Defining spot boundaries using regular and irregular feature shapes. A spot of 100 nM 3'-AlexaFluor532 labelled DNA IVT template visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. The spot has been defined with a regular circular feature (1A) or an irregular feature shape, adjusted to the shape of the spot (1B).

The DNA IVT template spot size and shape are influenced by the volume of DNA IVT template solutions spotted. DNA IVT template spots that have been spotted using an automated arrayer robot usually have a relatively uniform shape and size and it can be tempting to define the spot boundary using a regular feature shape. However, using the regular feature shape assumes that an identical volume has been spotted at each spot location and that an identical boundary (shape and size) can be defined for each spot. Consequently, if spotting leads to variable spot volumes, non-uniform spot may lead to

inaccurate FI statistical analysis using a regular feature shape to define the spot boundary (Figure 4.6). For example, in the case of a smaller, misshapen spot, when the spot boundary is defined using a regular circle feature shape, significant background FI is included in the mean FI calculation (Figure 4.6, Spot 3). On the other hand, when using an irregular feature shape to define the spot boundary, the mean FI output is more representative of the actual volume of DNA IVT template that was spotted and the immobilised level of DNA IVT template on the array. For this reason, in this thesis irregular feature shapes have been used to define the spot boundaries of the DNA IVT template array spots.

Regular Circle Feature (DNA Template 200 nM Alexa 532)			
Measurement	**Spot 1**	**Spot 2**	**Spot 3**
Fluorescence Mean	15104	16725	392
Background Mean	216	365	218
Fluorescence - Background	14888	16360	**174**



Irregular Shape Feature (DNA Template 200 nM Alexa 532)			
Measurement	**Spot 1**	**Spot 2**	**Spot 3**
Fluorescence Mean	15591	16998	11147
Background Mean	218	354	226
Fluorescence - Background	15373	16644	**10921**

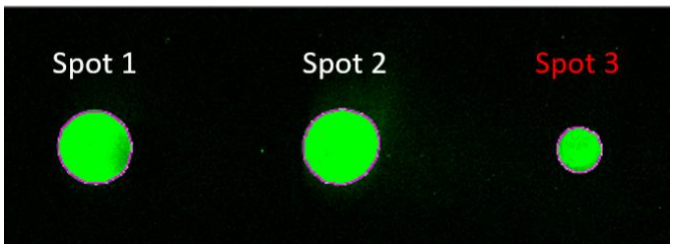


Figure 4.6. Defining spot boundaries using circular and irregular feature alignments for DNA. Three spots of 200 nM 3'-AlexaFluor532 labelled *ilvI* DNA IVT template visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. Spot 3 is visibly smaller than Spot 1 or 2, indicating that a smaller volume may have been spotted at the location of Spot 3. This may be the result of human error e.g., when manually spotting samples. (Upper panel) FI statistical analysis for the three spots using a regular circle feature shape to define the spot boundaries. Using the regular circle feature shape to define the spot boundary for Spot 3 incorrectly includes a significant amount of background FI in the mean FI calculation. (Lower panel) FI statistical analysis for the three spots using an irregular feature shape to define the spot/feature boundaries.

4.3. Generation of RNA arrays from a DNA IVT template array

4.3.1. Introduction

Following the production of the DNA IVT template array, the next step was to use this to generate the corresponding RNA array using the standard RNA array method (Figure 4.7A; Chapter 2 (section 2.2); Phillips *et al*., 2018, Henderson *et al*., 2019; Vincent *et al*., 2021 *in press*, Appendix 15). A DNA IVT template array slide – IVT mixture – SA-coated RNA capture slide "sandwich" was assembled. As IVT proceeds, RNA synthesised from each of the DNA IVT templates was captured *in situ* on an SA-coated RNA capture slide, via its 3′ SA aptamer, to generate the corresponding RNA array. It is possible to label the RNA *in situ* by supplementing the IVT mixture with Cy-labelled UTP (Phillips *et al*., 2018). Alternatively, the levels of unlabelled RNA can be visualised and quantified by probing the RNA array, post-array production, with a fluorescently labelled antisense ssDNA probe complementary to the SA linker region between the mRNA and the SA aptamer (Norouzi *et al*., 2019; Henderson *et al*., 2019; Vincent *et al*., 2021 *in press*). In this thesis, RNA arrays were probed post-array production with an antisense ssDNA probe (Figure 4.7B).

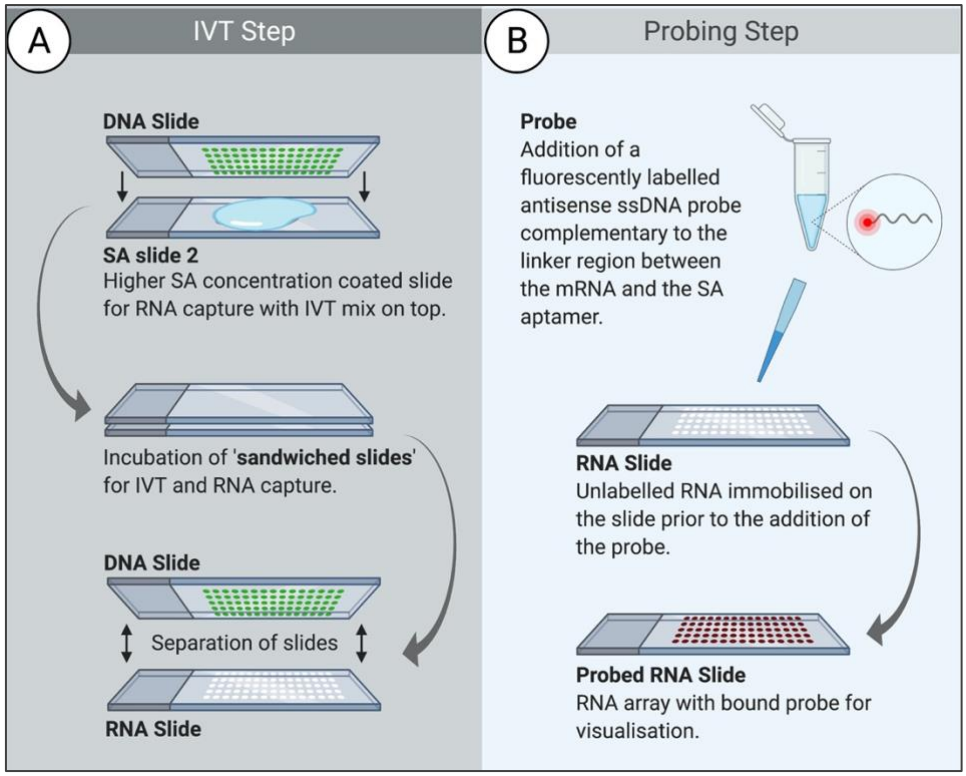


Figure 4.7. Generation of an RNA array. (A) Generation of an RNA array by IVT and *in situ* RNA capture. A DNA IVT template array – IVT mixture – SA-coated RNA capture surface "sandwich" is assembled. The surface of the DNA IVT template array slide with the spots of immobilised DNA IVT template and the SA-coated surface of the RNA capture slide both face inwards. The IVT mixture is pipetted between the DNA IVT template array slide and the RNA capture slide. Since the immobilised DNA IVT templates are in close proximity to the RNA capture surface, during IVT, the nascent RNA molecules directly immobilise to the SA-coated surface on the opposite slide, via the 3′ SA aptamer, with limited lateral diffusion. The slides are then separated and washed in 1 x Phosphate-Buffered Saline supplemented with 0.02% Tween 20 (PBST). **(B)** The unlabelled immobilised RNA on the RNA array slide may be probed with a fluorescently labelled antisense ssDNA probe complementary to the linker region between the mRNA and the SA aptamer.

4.3.2. IVT and *in situ* capture of the *ilvI* mRNA target

100 nM 5′-biotinylated and 3′-Alexa532 labelled *ilvI* DNA IVT template was spotted onto a SA-coated slide to generate a DNA IVT template array (Figure 4.8, 1). This DNA IVT template array was used to generate an *ilvI* RNA array by IVT and *in situ* capture. The *ilvI* RNA array was probed with the 5′-Alexa647-GTGTGTGTGTGTGTGTGTGT-3′ antisense ssDNA probe (SA-linker probe), complementary to the linker between the ROI (*ilvI*) and the SA aptamer (Figure 4.8, 2). The SA-linker probe was designed in-house and synthesised by Invitrogen (Thermo Fisher Scientific).

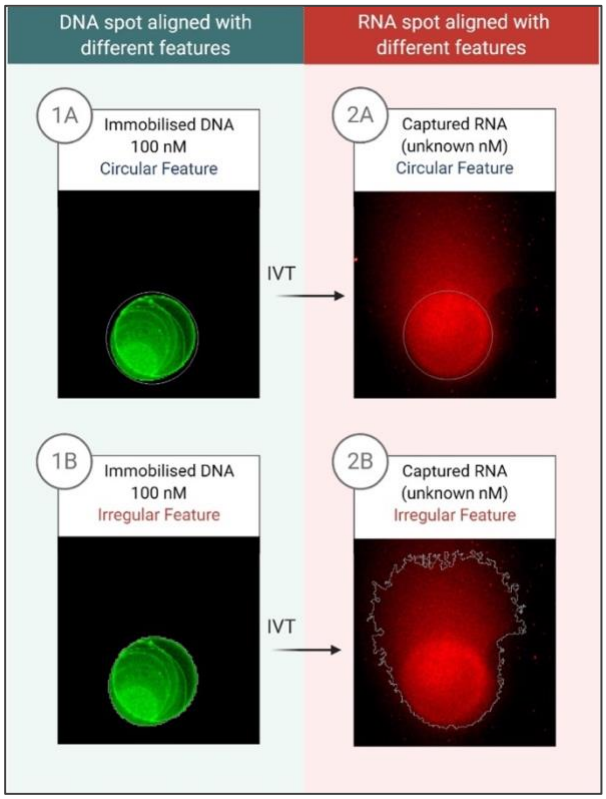


Figure 4.8. IVT and *in situ* capture of the *ilvI* mRNA target. (1) 100 nM of 5′-biotinylated and 3′-Alexa532 labelled *ilvI* DNA IVT template visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. The spot boundary was defined using a regular circle feature shape (1A) or an irregular feature shape (1B). (2) *ilvI* RNA generated by IVT and *in situ* capture of the DNA IVT template in (1), probed with 5′-Alexa647-GTGTGTGTGTGTGTGTGTGT-3′ antisense ssDNA (SA-linker probe) and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. The spot boundary was defined using a regular circle feature shape (2A) or an irregular feature shape (2B).

4.3.2.1. Defining the array spot boundaries with regular and irregular spot shapes

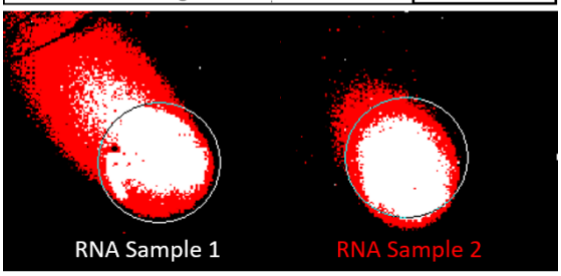
The boundaries of the DNA IVT template spots and the SA-linker probe bound to the immobilised *ilvI* mRNA target spots were defined in the integrated scanner software using a regular circle feature shape (Figure 4.8, 1A and 2A) and an irregular feature shape (Figure 4.8, 2A and 2B). Although the DNA IVT template spots are often consistent in size and shape, as discussed above, the size and shape of the RNA spot is less uniform due to diffusion during IVT and *in situ* capture (compare Figure 4.8, 1 and 4.8, 2). Transcribed and immobilised RNA has clearly diffused outside the boundaries set by a regular circle feature shape based on the corresponding DNA IVT template spot (Figure 4.8, 2A). This would lead to the incorrect exclusion of FI when calculating the mean FI of the spot/feature and also

the incorrect inclusion of FI when calculating the mean background FI. However, this can be avoided by using the irregular feature alignment (Figure 4.8, 2B).

4.3.2.2. Regular vs irregular alignment features of RNA spots

The choice of feature shape for the RNA spots is particularly important when comparing two or more RNA spots because the RNA spots may differ in shape and size due to different relative transcription efficiencies. For example, a DNA IVT template with a high transcription efficiency may produce more RNA, more quickly allowing diffusion over a wider area during IVT and *in situ* capture, than a DNA IVT template with a low transcription efficiency. When using a regular circle feature shape to define the boundary of these RNA spots, more FI will be excluded for the RNA that is transcribed with high efficiency than for the RNA that is transcribed with lower efficiency (Figure 4.9). This will underestimate the RNA yield for the RNA transcribed with high efficiency. In contrast, when using the irregular feature shape to define the spot boundaries, the total FI, and therefore total RNA yield, is calculated for both RNAs. For the analysis presented in this thesis, the irregular feature shape was used to define the RNA spot boundaries and the mean FI minus the mean background FI was calculated.

Regular Circle Feature (RNA Probed Alexa 647)		
Measurement	**RNA Sample 1**	**RNA Sample 2**
Fluorescence Mean	65535	65535
Background Mean	21454	21079
Fluorescence - Background	44081	**44456**



Irregular Shape Feature (RNA Probed Alexa 647)		
Measurement	**RNA Sample 1**	**RNA Sample 2**
Fluorescence Mean	57647	42951
Background Mean	14602	11232
Fluorescence - Background	43045	**31719**

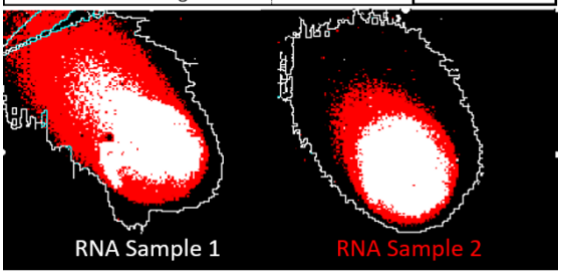


Figure 4.9. Defining RNA spot boundaries using regular and irregular feature shape. Two RNA spots probed with SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. (Left) FI statistical analysis for the two RNA spots using a regular circle feature shape to define the spot/feature boundaries. (Right) FI statistical analysis for the two RNA spots using an irregular feature shape to define the spot/feature boundaries.

4.4. Generating RNA arrays of APP GcvB putative mRNA targets

4.4.1. Introduction

Having designed the strategy for generating, visualising, and quantifying an RNA array of putative mRNA targets of GcvB, the next step was to generate the mRNA target array *in vitro*. In this section, a preliminary experiment is presented in which a low-density array of the DNA IVT templates for each of the mRNA targets was prepared by manual spotting. This was then used to generate a low-density RNA array of the mRNA targets. The aim of this experiment was to provide a yes-no output as to whether the DNA IVT templates immobilised to the DNA IVT template array slide were transcribed, and the RNA produce captured by the RNA capture slide.

4.4.2. Generating a low-density array of putative mRNA targets of GcvB

A sample (0.2 µl) of 200 nM of each 5′-biotinylated and 3′-Alexa647 labelled DNA IVT template was spotted, in duplicate, onto an SA-coated slide by manual spotting to generate the mRNA target DNA IVT template array. This DNA IVT template array was then used to generate an unlabelled mRNA target array following the standard RNA array protocol (Chapter 2, section 2.2.5). The mRNA target array was then probed with 500 nM SA-linker probe. A schematic of the experimental design is shown in Figure 4.10. The aim of this experiment was to provide a yes-no output as to whether the DNA IVT templates immobilised to the DNA IVT template array slide were transcribed and captured by the RNA capture slide.

In order to facilitate RNA immobilisation on the RNA capture slide, the mRNA target RNA, including the SA aptamer, must be fully transcribed and folded correctly. The secondary structure(s) of each mRNA target were predicted using RNAFold and suggested that the conjugated SA aptamer was likely to fold into the correct structure in each case (for each mRNA target) (Appendix 10). Furthermore, the 3′ location of the SA aptamer in

the RNA construct for each mRNA target ensures that only fully transcribed RNA molecules should be immobilised on the RNA capture slide. To confirm these predictions and validate the experimental strategy, it is necessary to check that the DNA IVT templates are transcribed into full-length RNA products and captured by the SA-coated RNA capture slide.

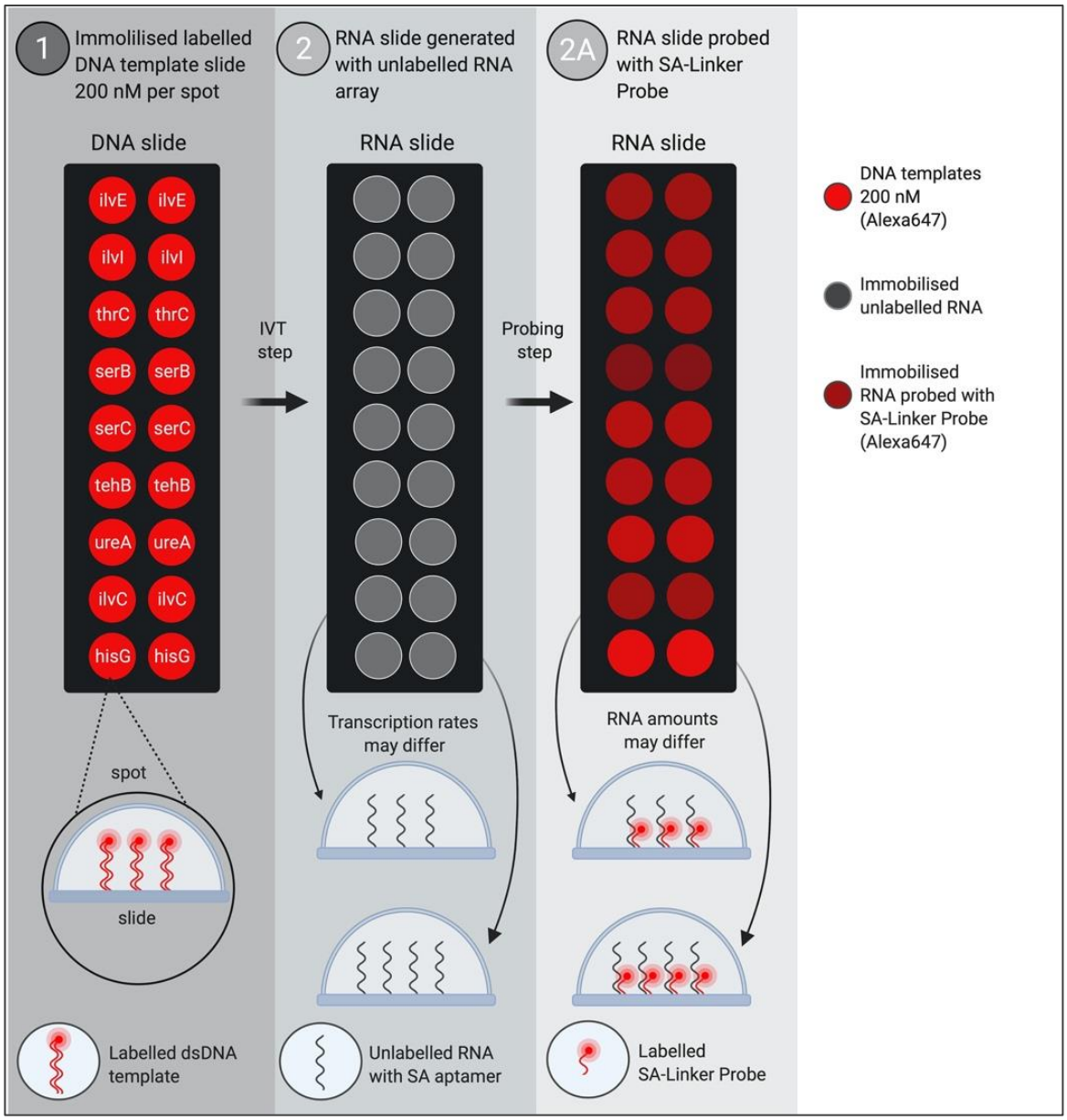


Figure 4.10. Schematic of the generation of a low-density RNA array of the putative mRNA targets of GcvB. (1) Samples (0.2 µl) of 200 nM of the 5′-biotinylated and 3′-Alexa647 labelled DNA IVT template for each of the predicted mRNA targets of GcvB manually spotted, in duplicate, on an SA-coated slide to produce a DNA IVT template array. The DNA IVT template array is visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. (2) A corresponding unlabelled mRNA target array is generated from the DNA IVT template array in (1) by IVT and *in situ* capture. (2A) The mRNA target array is probed with Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter.

As shown Figure 4.11, 1, each of the DNA IVT templates were successfully immobilised to the SA-coated slide to generate the DNA IVT template array. FI was only detected at locations where the DNA IVT template had been manually spotted, there was no non-specific binding to other parts of the slide. Similarly, all the mRNA targets were successfully transcribed and immobilised on the RNA capture surface as an RNA array (Figure 4.11, 2). Following probing of the mRNA target array with the SA-linker probe, FI was only detected in discrete spots indicating that the RNA was immobilised in sites correspondent to their DNA IVT template spots and that the SA-linker probe bound specifically to the immobilised RNA.

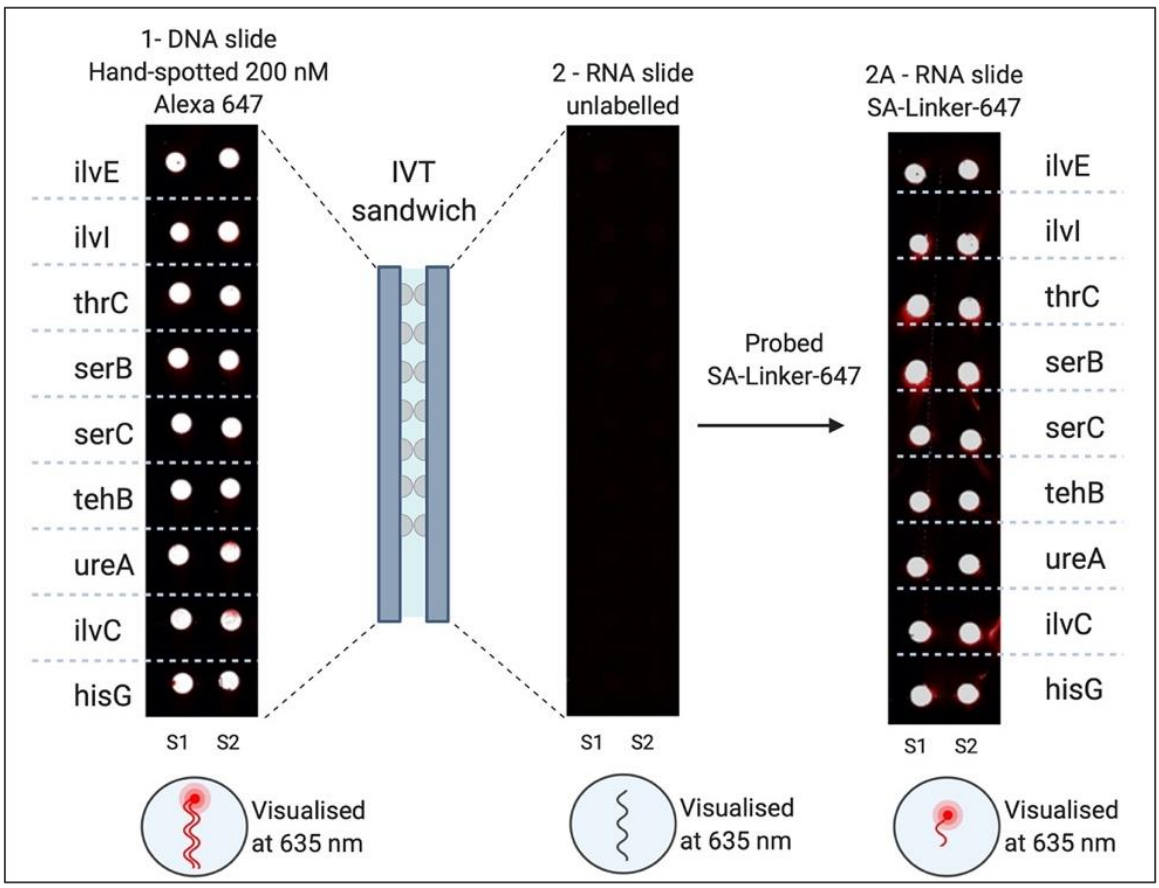


Figure 4.11. Generation of a low-density RNA array of putative mRNA targets of GcvB. (1) Samples (0.2 µl) of 200 nM of the 5′-biotinylated and 3′-Alexa647 labelled DNA IVT template for each of the predicted mRNA targets of GcvB was manually spotted, in duplicate, on an SA-coated slide to produce a DNA IVT template array. The DNA IVT template array was visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. (2) A corresponding unlabelled mRNA target array was generated from the DNA IVT template array in (1) by IVT and *in situ* capture. Visualisation using an excitation wavelength of 635 nm and a Standard Red emission filter revealed no FI. (2A) The mRNA target array was probed with Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter.

4.4.3. Generating high-density arrays of putative mRNA targets of GcvB

Having demonstrated that a low-density array of DNA IVT templates could be produced and used to generate a low-density RNA array of the putative mRNA targets of GcvB, the aim of the next set of experiments was to produce a high-density DNA IVT template array by automated spotting with an arrayer robot and use this to generate a high-density RNA array of mRNA targets. Moving to a high-density array format allows for rapid, high-throughput optimisation and analysis of the RNA array generation process e.g. optimisation of the concentration of the spotted DNA IVT template and analysis of transcription efficiencies.

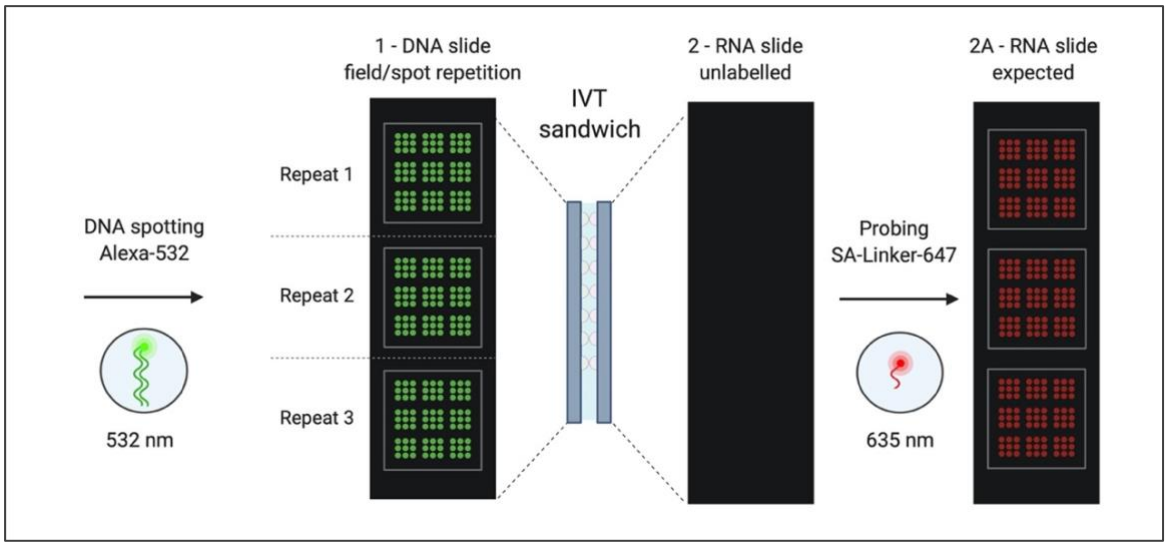


Figure 4.12. Schematic of the layout of high-density arrays. (1) 5′-biotinylated and 3′-Alexa532 labelled DNA IVT templates are spotted by an automated arrayer robot in three grids (Repeat 1, 2 and 3) of 3 x 3 fields of 3 x 3 spots (243 spots per array). Each grid typically contains one field for each of the putative mRNA targets of GcvB. The DNA IVT template array is visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. (2) A corresponding unlabelled mRNA target array is generated from the DNA IVT template array in (1) by IVT and *in situ* capture. (2A) The mRNA target array is probed with Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter.

The general array layout for the high-density array experiments is shown in Figure 4.12. Two sets of experiments were then planned. The first experiment involved spotting concentration gradients for each DNA IVT template to determine the transcription efficiency for each DNA IVT template (Figure 4.13, A). The second experiment involved

spotting each DNA IVT template multiple times at the fixed concentration of 200 nM to compare the relative transcription efficiencies between DNA IVT templates (Figure 4.13, B).

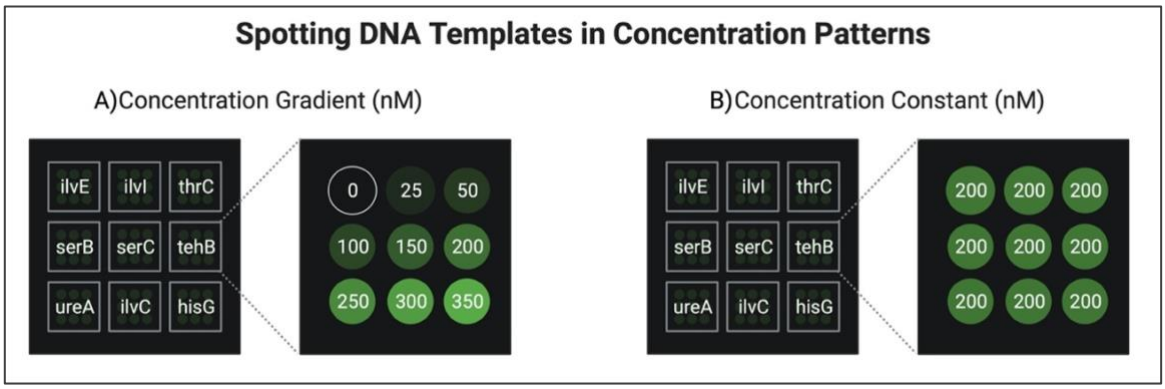


Figure 4.13. Schematics of the layout of grids and fields on high density DNA IVT template arrays. Each grid contains 3 x 3 fields with each field containing 3 x 3 spots of one of the DNA IVT templates. (A) A concentration gradient of DNA IVT template is spotted (0 nM, 25 nM, 50 nM, 100 nM, 150 nM, 200 nM, 250 nM, 300 nM and 350 nM). (B) A fixed concentration of 200 nM DNA IVT template is spotted multiple times.

4.4.3.1. Transcription efficiencies of individual DNA IVT templates

It is expected that for a given DNA IVT template the amount of RNA that is transcribed and captured *in situ* will be proportional to the amount of DNA IVT template immobilised. However, each DNA IVT template may be transcribed and captured with a different efficiency. To investigate this, each 5'-biotinylated and 3'-Alexa532 labelled DNA IVT template was spotted at a concentration of 0, 25 nM, 50 nM, 100 nM, 150 nM, 200 nM, 250 nM, 300 nM and 350 nM by an automated arrayer robot (Figure 4.14). The resulting DNA IVT template array was used to generate an unlabelled mRNA target array that was probed with 500 nM SA-linker probe.

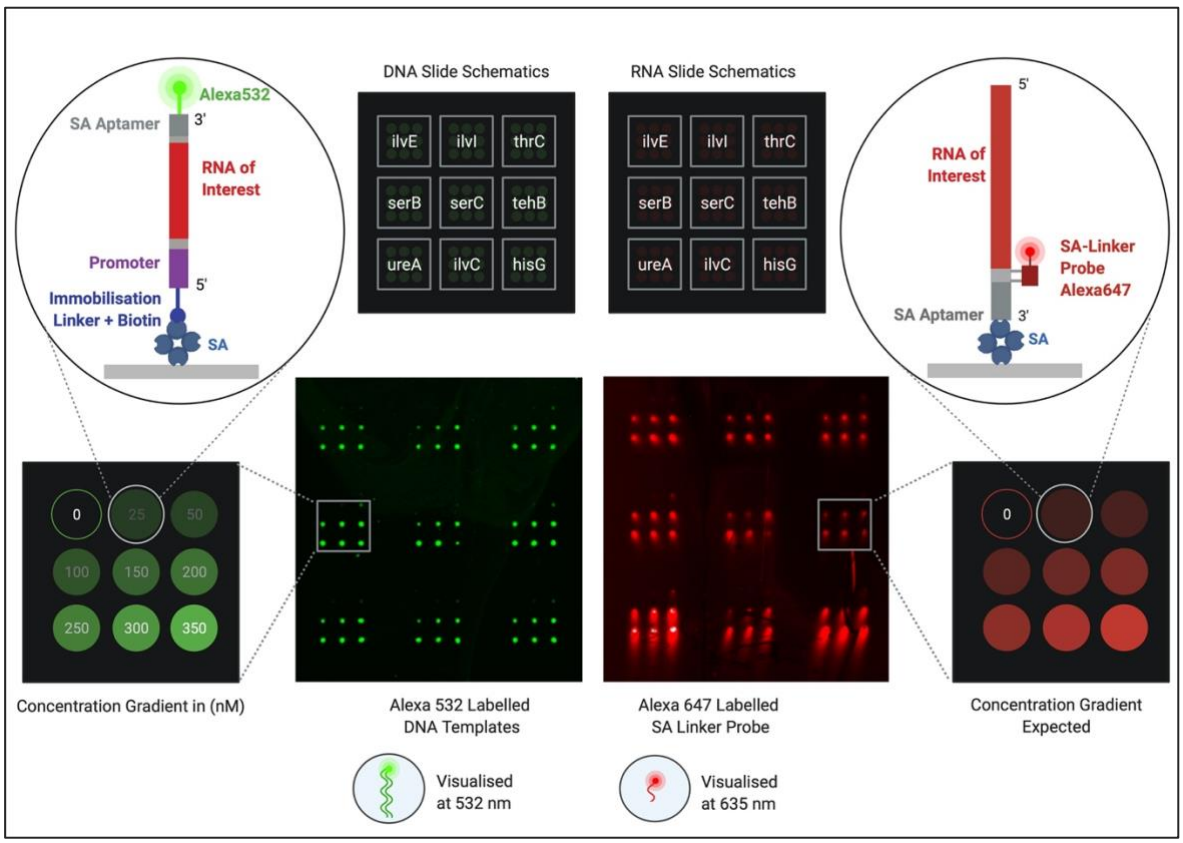


Figure 4.14. Spotting a concentration range of DNA IVT templates. (Left-hand side) A grid of 3 x 3 fields with each field containing 3 x 3 spots of one of the 5'-biotinylated and 3'-Alexa532 labelled DNA IVT templates spotted in a concentration gradient (0 nM, 25 nM, 50 nM, 100 nM, 150 nM, 200 nM, 250 nM, 300 nM and 350 nM) by an automated arrayer robot. The DNA IVT array was visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. (Right-hand side) An RNA array was generated by IVT and *in situ* capture, probed with 500 nM Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. Grids were repeated in triplicate on a single slide.

DNA spots were clearly visible for each of the DNA IVT templates spotted at a concentration of 50 nM and above. The size and shape of the DNA spots is relatively uniform. As expected, the DNA FI signal appears to increase with increasing concentration of spotted DNA IVT template for each of the templates. Similarly, RNA spots could be detected at locations corresponding to DNA IVT template concentrations of 50 nM and above for all each of the DNA IVT templates, confirming that the mRNA targets had been transcribed and captured on the RNA capture slide. In the *ilvE* and *ureA* fields, a faint RNA spot is detected in the location corresponding to 25 nM DNA IVT template spot, implying that a low level of DNA IVT template was immobilised when spotted at a concentration 25 nM. The size and shape of the RNA spots is more variable than for the DNA spots, presumably due to variable diffusion of the mRNA target, and may reflect variable

transcription efficiencies. For example, spots in the *ureA* and *hisG* fields clearly show RNA diffusion out of the corresponding DNA IVT template spot boundaries, suggesting that these DNA templates may have a higher transcription efficiency than the other mRNA targets. For each of the mRNA targets, the levels of RNA appear to increase with increasing concentration of spotted DNA IVT template.

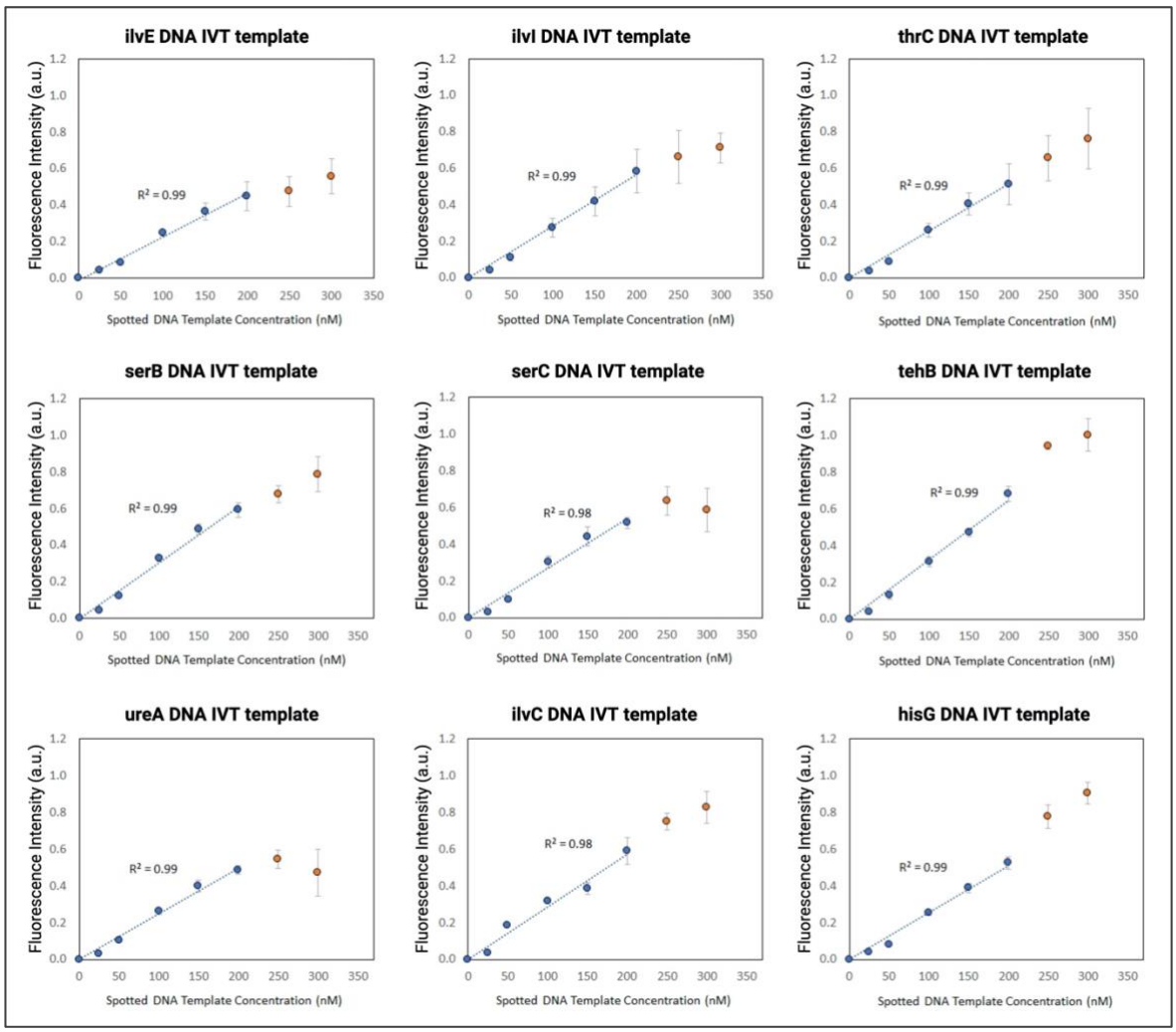


Figure 4.15. Plot of mean DNA spot FI (a.u.) against concentration of spotted DNA IVT template (nM). Data have been normalised to the DNA spot with the highest FI using the min-max normalisation. Data are the mean of three experimental repeats and error bars represent the standard error of the mean. Data are fit to a linear equation. Data points within the linear range are coloured blue and data points outside the linear range are coloured red.

The mean FI was calculated for each of the DNA IVT template spots and plotted against the spotted concentration (Figure 4.15). The mean FI for the DNA IVT template spots increases linearly with increasing spotted concentration, up to a spotted concentration of 200 nM, for all templates. At concentrations of spotted IVT DNA template above 200 nM the FI of some is non-linear. This might suggest that the spot location becomes saturated with DNA IVT template molecules when spotted at concentrations over 200 nM and that the capacity of the surface has been reached at this concentration. Therefore, DNA IVT template concentrations up to 200 nM may be a more suitable working range for subsequent experiments.

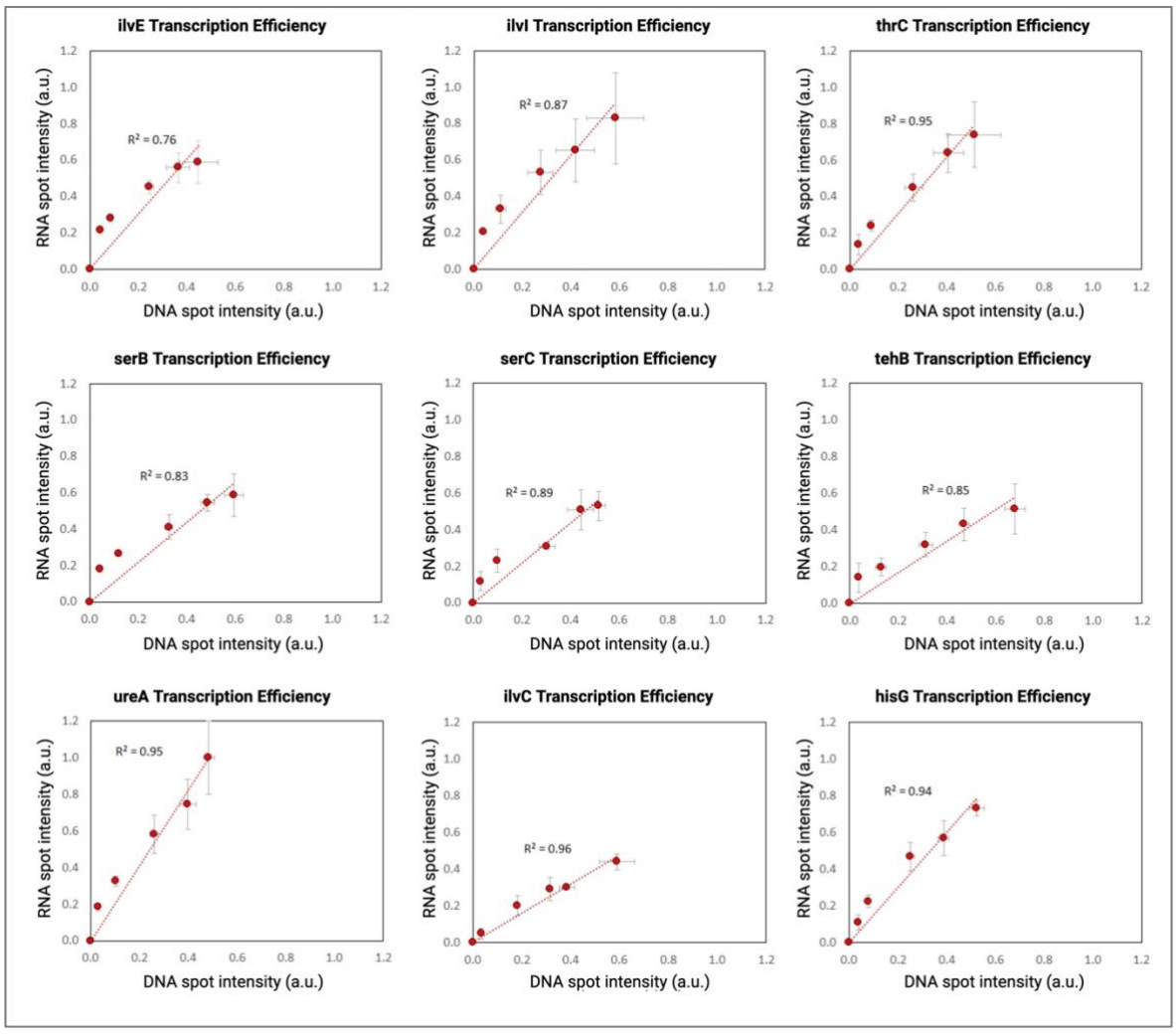


Figure 4.16. Plot of RNA spot FI (a.u.) against DNA spot FI (a.u.). Data have been normalised to the RNA spot with the highest FI using the min-max normalisation. Data are the mean of three experimental repeats and error bars represent the standard error of the mean. Only data points corresponding to the DNA FI linear range are plotted. Data are fit to a linear equation.

The mean FI was calculated for SA-linker probe bound to each of the mRNA target spots and plotted against the mean FI for the corresponding DNA IVT template spot for DNA IVT template spots in the DNA spot FI linear range (Figure 4.16). Similar to the results for the DNA IVT template FI, RNA spot fluorescence increases with increasing DNA IVT template spot FI. A moderate linear correlation is observed between DNA and RNA spot FI. The maximum RNA levels varied from 0.2 a.u. for *tehB* and *ilvI* up to 0.6 a.u. for *hisG and* 1 a.u. for *ureA*. This suggests variable transcription efficiencies between the DNA IVT templates with low transcription efficiencies for *tehB* and *ilvI* and higher transcription efficiencies for *hisG* and *ureA*.

4.4.3.2. Transcription efficiencies between DNA IVT templates

The data presented in the previous section suggest that there are differences in the transcription efficiency for the different DNA IVT templates. To investigate this further it was decided to compare the transcription efficiency for each DNA IVT template by spotting each 5′-biotinylated and 3′-Alexa532 labelled DNA IVT template 27 times (three 3 x 3 fields), on a single array, at a fixed concentration of 200 nM (Figure 4.17). A concentration of 200 nM DNA IVT template was chosen because this generated the highest RNA yield within the linear range for each of the templates (Figure 4.16). The DNA IVT template array was used to generate the corresponding mRNA target array, and this was probed with 500 nM SA-linker probe.

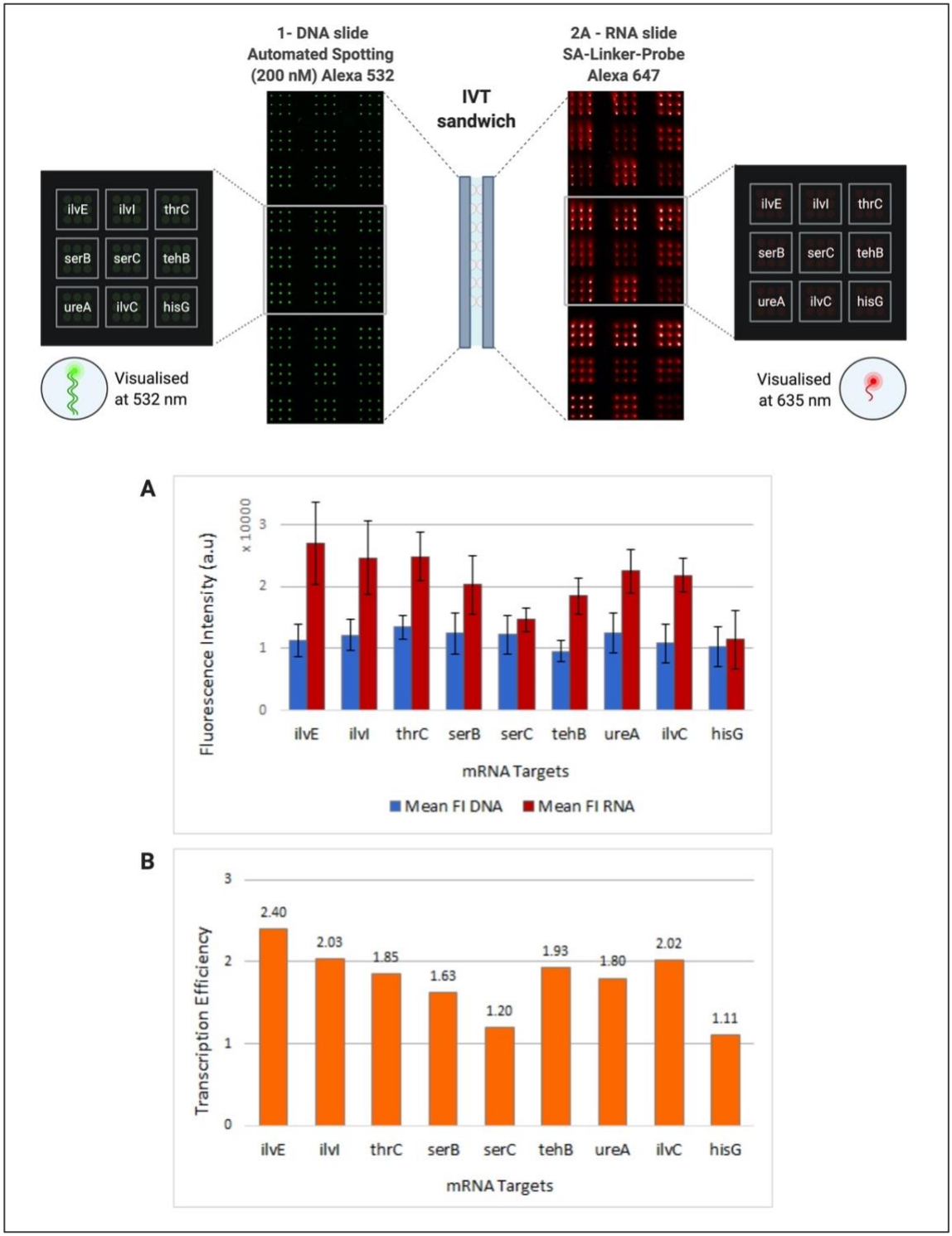


Figure 4.17. Spotting DNA IVT templates at a fixed concentration of 200 nM. (1) A grid of 3 x 3 fields with each field containing 3 x 3 spots of one of the 5′-biotinylated and 3′-Alexa532 labelled DNA IVT templates spotted at a fixed concentration of 200 nM using an automated arrayer robot. The DNA IVT array was visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. (2A) An RNA array was generated by IVT and *in situ* capture, probed with 500 nM Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. Grids were repeated in triplicate on a single slide. (A) A plot of mean FI (in a.u.) for the DNA IVT template spots (blue bars) and the SA-linker probe bound to the corresponding mRNA target spots (red bars). Data are the mean of 27 spots and error bars represent the standard error of the mean. (B) The transcription efficiencies were calculated by normalising the mean FI for the SA-linker bound to the mRNA target spots to the mean FI for the corresponding DNA IVT template spots. For example, for *ilvE*, mRNA Fl of 2.7/1.1 Fl for DNA gives a transcription efficiency for ilvE of 2.4.

DNA spots were clearly visible for each of the DNA IVT templates spotted and they appear relatively uniform. Corresponding RNA spots are also visible for each of the DNA spots indicating that each DNA IVT template spot was transcribed, and the RNA captured. For a given DNA IVT template, the resultant RNA spots appear relatively uniform. However, there are clear differences in the RNA levels between different mRNA targets. For example, the *ilvE* and *ilvC* mRNA target levels appear to be consistently high and the *serC* and *hisG* mRNA target levels appear to be consistently low. This confirms the previous data suggesting that DNA IVT templates are transcribed with varying efficiencies.

The mean FI was calculated for each of the DNA IVT template spots and for the SA-linker bound to each of the mRNA target spots (Figure 4.17, A). As expected, given that each of the DNA IVT templates were spotted at the fixed concentration of 200 nM, the FI for the DNA spots is similar for each of the DNA IVT templates. In contrast, the FI for the SA-linker probe bound to the mRNA target spots is highly dependent on the mRNA target. Normalising the mean FI for the SA-linker bound to the mRNA target spots to the mean FI for the DNA IVT template (Figure 4.17, B) confirms that the levels of the immobilised mRNA targets differ, most likely due to different transcription efficiencies.

4.5. Application of RNA arrays to the validation of putative sRNA-mRNA interactions

Once the conditions for generating RNA arrays of putative mRNA targets of GcvB had been established, the next step was to investigate whether GcvB could interact with the target mRNAs on an RNA array. RNA interaction studies can be carried out post-RNA array production, by adding the interacting partner in solution (Phillips *et al*., 2018, Henderson *et al*., 2019), or *in situ*, by including a DNA IVT template encoding the interacting partner (Norouzi *et al.,* 2019) or by supplementing the IVT mixture with the interacting partner (Henderson *et al*., 2019). Here, two strategies were tested. The first strategy involved probing an RNA array of the putative mRNA targets of GcvB post-array production with a solution of fluorescently labelled GcvB. The second strategy involved co-spotting the DNA IVT template for the putative mRNA target together with the DNA IVT template for GcvB, so that transcription of both the mRNA target and the sRNA occurs simultaneously. This allows GcvB to bind to the mRNA targets *in situ*, during the IVT/capture step of the RNA array protocol. The bound GcvB is then probed post-array production with a fluorescently labelled antisense ssDNA probe. Both strategies are detailed below.

4.5.1. Probing RNA arrays of the putative mRNA targets of GcvB post-array production with fluorescently labelled GcvB

The initial strategy to investigate the interaction between GcvB and the putative mRNA targets was to probe the mRNA target array post-array production with a fluorescently labelled GcvB (GcvB-Cy3) (Figure 4.18). GcvB-Cy3 was GcvB, internally labelled with Cy3, and was synthesised by IVT using an IVT mixture supplemented with Cy3-labelled UTP (Chapter 2, section 2.2.3.5 and 2.2.7.1), with random but proportional incorporation. Cy3 fluorescence will only be detected for spots containing an mRNA target to which GcvB-Cy3 is bound. The presence of the mRNA targets on the array is confirmed for each position by probing with SA-linker probe. GcvB-Cy3 and SA-linker probe (Alexa647 labelled) contain orthogonal fluorophores which can be detected simultaneously.

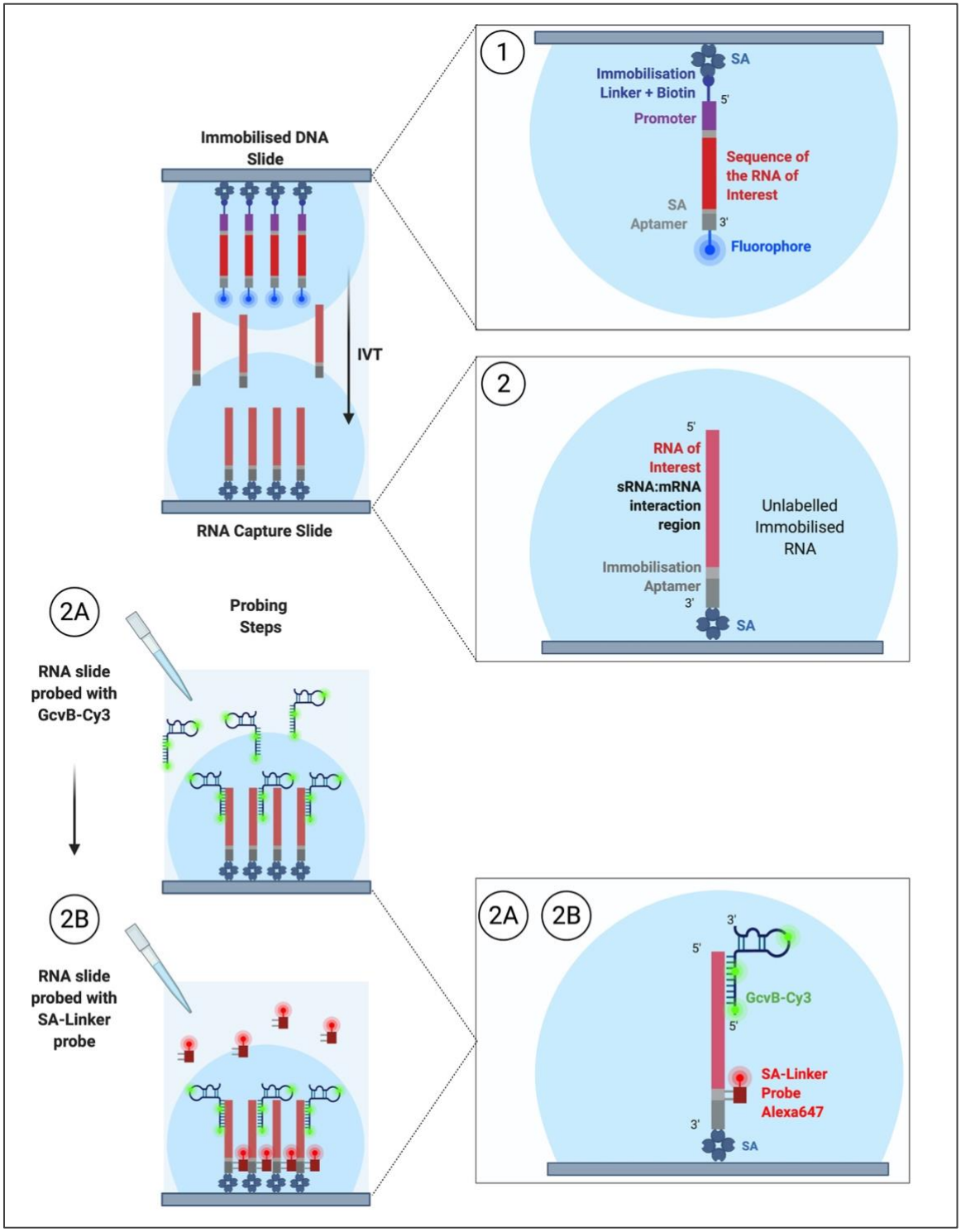


Figure 4.18. Schematic of an mRNA target array probed with internally labelled GcvB-Cy3. (1) An SA-coated slide is spotted with 5′-biotinylated, 3′-Alexa488/Alexa647 labelled DNA IVT templates encoding the putative mRNA targets of GcvB to generate the DNA IVT template array. (2) An mRNA target array is generated from the DNA IVT template array by IVT and *in situ* RNA capture. (2A) The mRNA target array is probed post-array production with GcvB-Cy3 to screen for GcvB-mRNA target interactions, (2B) and then with SA-linker probe, to enable visualisation and quantification of the levels of mRNA target on the RNA array. The probed array is visualised using an excitation wavelength of 532 nm and a Standard Green emission filter and using an excitation wavelength of 635 nm and a Standard Red emission filter.

4.5.1.1. Probing low-density RNA arrays of putative mRNA targets of GcvB post-array production with GcvB-Cy3

Samples (0.2 µl) of 200 nM of each 5'-biotinylated and 3'-Alexa647 labelled DNA IVT template were spotted, in duplicate, onto a SA-coated slide by manual spotting to generate the mRNA target DNA IVT template array. This DNA IVT template array was then used to generate an unlabelled mRNA target array. The mRNA target array was then probed with 500 nM of GcvB-Cy3, to screen for GcvB-mRNA target interactions. Finally, the mRNA target array was probed with 500 nM SA-linker probe, to visualise and quantify the RNA levels. A schematic of the experimental design is shown in Figure 4.19.

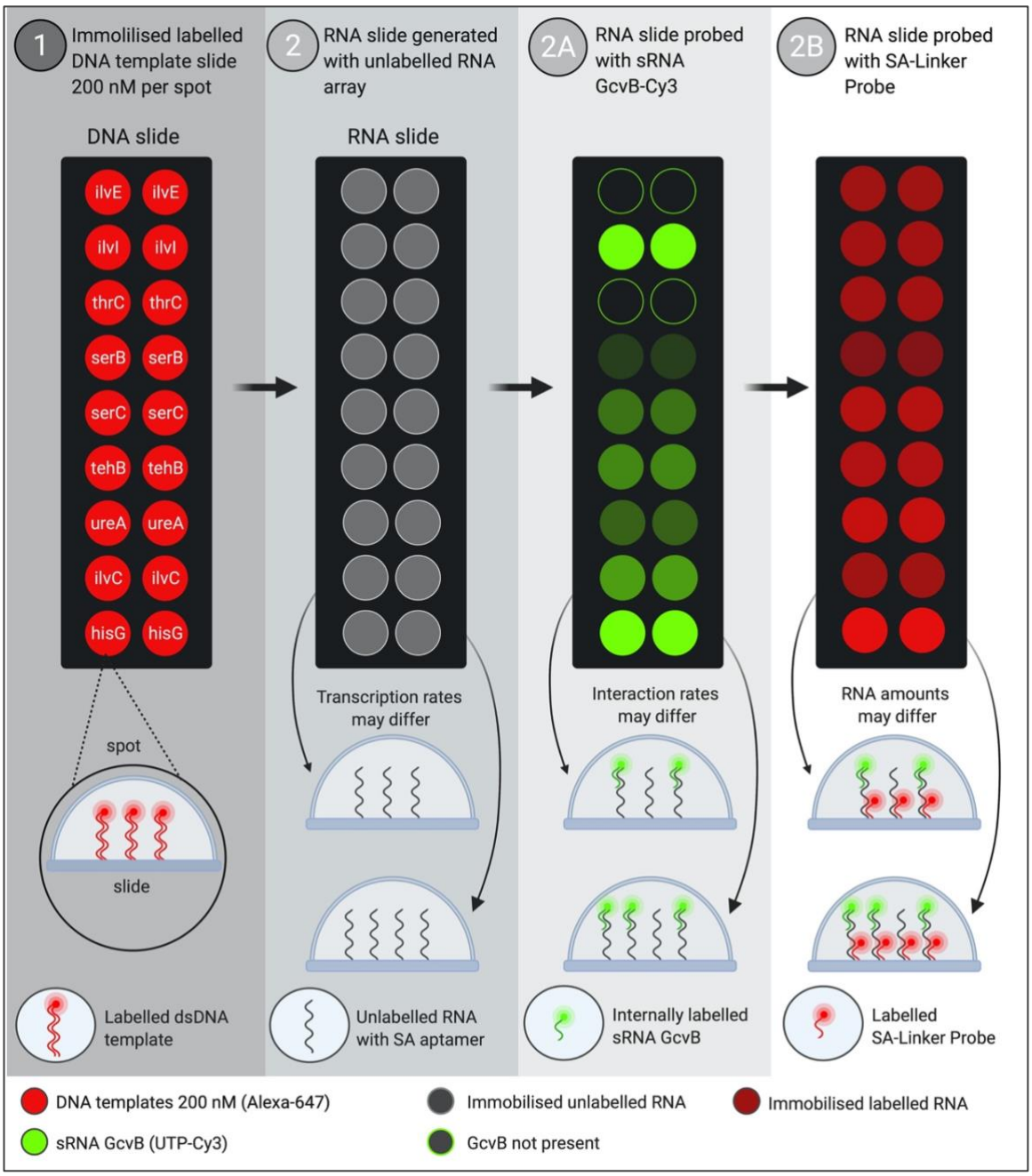


Figure 4.19. Schematic of the generation and probing of a low-density RNA array of putative mRNA targets of GcvB. (1) A sample (0.2 µl) of 200 nM of the 5'-biotinylated and 3'-Alexa647 labelled DNA IVT template for each of the predicted mRNA targets of GcvB was manually spotted, in duplicate, on an SA-coated slide to produce a DNA IVT template array. The DNA IVT template array is visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. (2) A corresponding unlabelled mRNA target array is generated from the DNA IVT template array in (1) by IVT and *in situ* capture. (2A) The mRNA target array is probed with GcvB-Cy3 and visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. (2B) The mRNA target array is probed with Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter.

As shown Figure 4.20, 1, each of the DNA IVT templates were successfully immobilised to the SA-coated slide to generate the DNA IVT template array. Prior to

probing, no FI was detected for the unlabelled RNA slide (Figure 4.20, 2). Following probing of the mRNA target array with GcvB-Cy3, significant FI was detected for spots containing the *ilvI, ilvC* and *hisG* mRNA targets (Figure 4.20, 2A), indicating that GcvB-Cy3 had bound to these mRNA targets on the RNA array. Weaker FI was detected for spots containing the remaining mRNA targets. To confirm that these mRNA targets were present on the array, and that the apparent lack of GcvB-Cy3 binding was not due to low levels of RNA, the mRNA target array was probed with SA-linker probe (Figure 4.20, 2B). Following probing of the mRNA target array with the SA-linker, FI was detected at each of the positions expected to contain immobilised RNA. Therefore, dimmed Cy3 fluorescence suggests that GcvB binds significantly less to the remaining mRNAs.

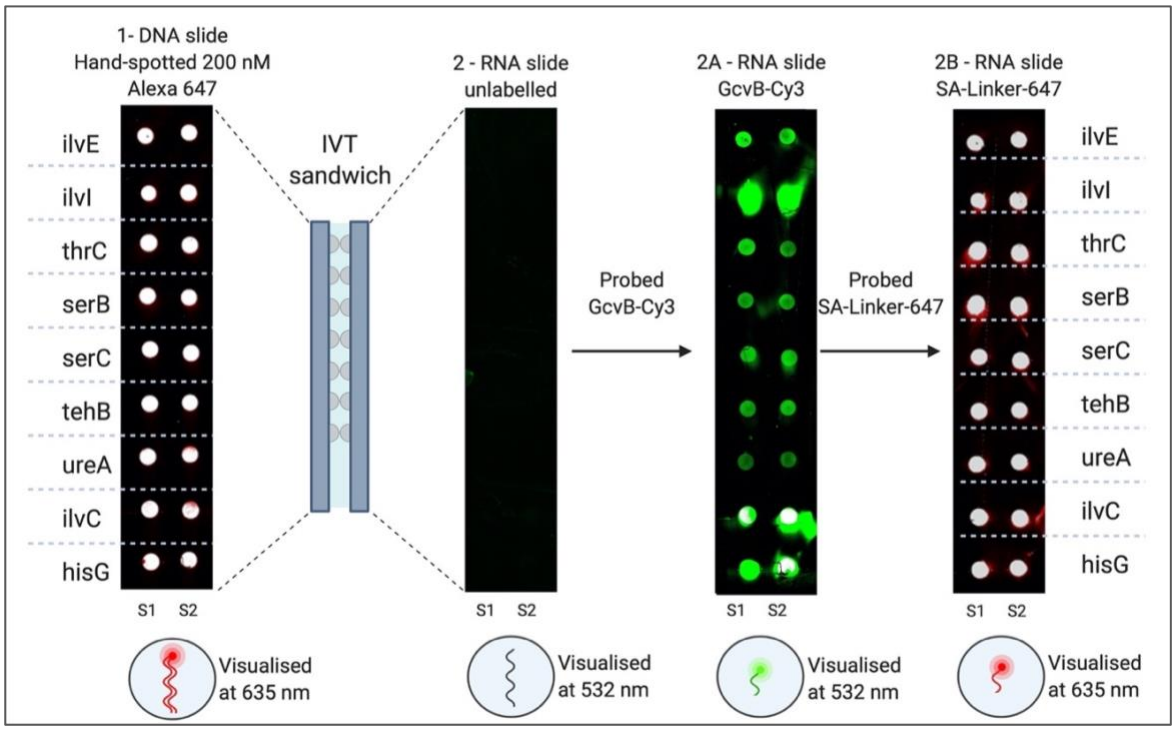


Figure 4.20. Generation and probing of a low-density RNA array of putative mRNA targets for GcvB. (1) A sample (0.2 µl) of 200 nM of the 5′-biotinylated and 3′-Alexa647 labelled DNA IVT template for each of the predicted mRNA targets of GcvB was manually spotted, in duplicate, on an SA-coated slide to produce a DNA IVT template array. The DNA IVT template array was visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. (2) A corresponding unlabelled mRNA target array was generated from the DNA IVT template array in (1) by IVT and *in situ* capture. Visualisation using an excitation wavelength of 532 nm and a Standard Green emission filter revealed no FI. (2A) The mRNA target array was probed with 500 nM GcvB-Cy3 and visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. (2B) The mRNA target array was probed with 500 nM Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter.

4.5.1.2. Probing high-density RNA arrays of putative mRNA targets of GcvB post-array production with GcvB-Cy3

Having demonstrated that GcvB-Cy3 binds to the potential mRNA targets on a low-density mRNA target array, the next step was to investigate this further using high-density RNA arrays. Moving to a high-density array format allows for rapid, high-throughput analysis of the GcvB-mRNA target interactions. For the high-density array experiments, each DNA IVT template was spotted multiple times at the fixed concentration of 200 nM in the general array layout shown in Figure 4.21.

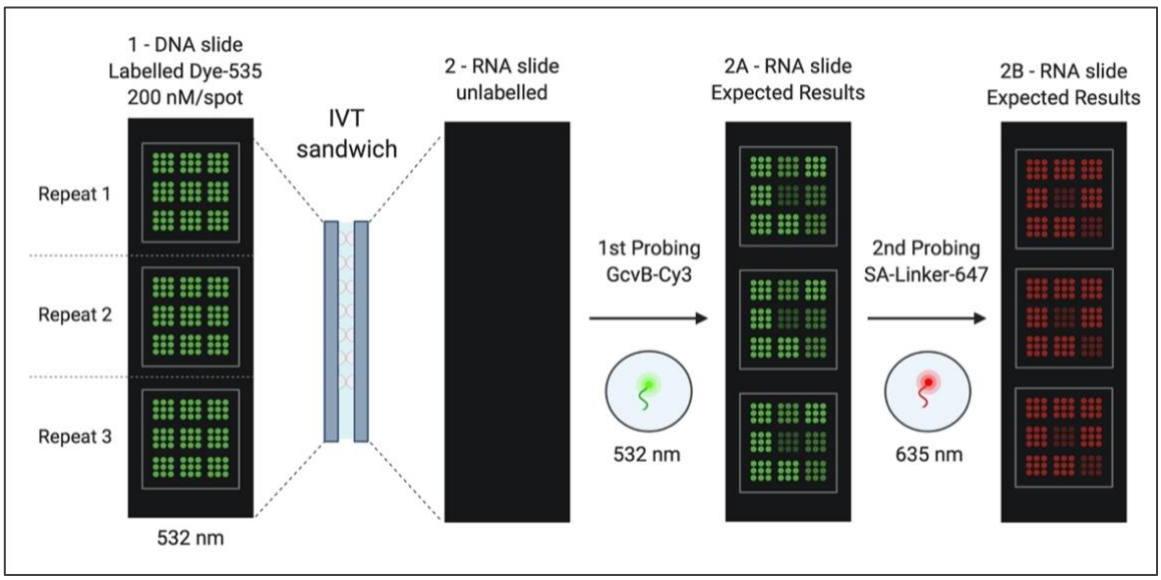


Figure 4.21. Schematic of the generation and probing of a high-density RNA array of putative mRNA targets of GcvB. (1) 5′-biotinylated and 3′-Alexa532 labelled DNA IVT templates are spotted by an automated arrayer robot in three grids (Repeat 1, 2 and 3) of 3 x 3 fields of 3 x 3 spots (243 spots per array). Each grid typically contains one field for each of the putative mRNA targets of GcvB. The DNA IVT template array is visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. (2) A corresponding unlabelled mRNA target array is generated from the DNA IVT template array in (1) by IVT and *in situ* capture. (2A) The mRNA target array is probed with GcvB-Cy3 and visualised using an excitation wavelength of 532 nm and a Standard Green emission filter. (2B) The mRNA target array is probed with Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter.

Each 5′-biotinylated and 3′-Alexa532 labelled DNA IVT template was spotted 27 times (three 3 x 3 fields), on a single array, at a fixed concentration of 200 nM to generate a DNA IVT template array. The DNA IVT template array was used to generate the

corresponding mRNA target array, and this was probed with 500 nM GcvB-Cy3 and 500 nM SA-linker probe (Figure 4.22).

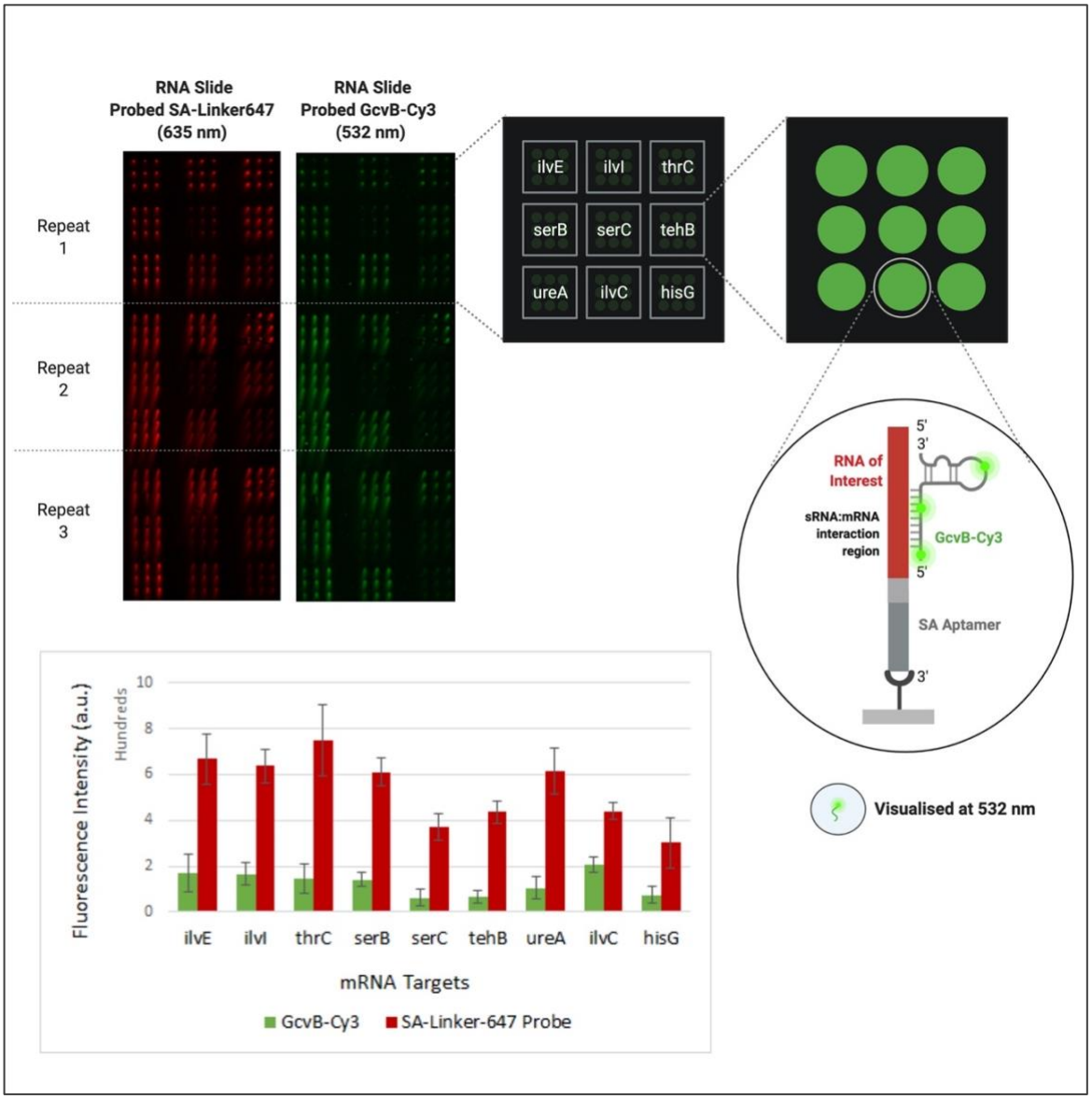


Figure 4.22. Generation and probing of a high-density RNA array of putative mRNA targets of GcvB. Three grids of 3 x 3 fields with each field containing 3 x 3 spots of one of the 5′-biotinylated and 3′-Alexa532 labelled DNA IVT templates were spotted at a fixed concentration of 200 nM using an automated arrayer robot. An RNA array was generated by IVT and *in situ* capture. The RNA array was probed with 500 nM GcvB-Cy3 and visualised using an excitation wavelength of 532 nM and a Standard Green emission filter. A schematic of GcvB-Cy3 bound to an mRNA target on the RNA array is shown. The RNA array was subsequently probed with 500 nM Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. A plot of mean FI (in a.u.) for GcvB-Cy3 bound to the RNA spots (green bars) and for SA-linker probe bound to the same spot (red bars).

SA-linker probe FI was detected for each of the mRNA target spots indicating that each DNA IVT template spot was transcribed, and the RNA captured *in situ*. For a given DNA

IVT template, the resultant RNA spots appear relatively uniform. However, as observed previously, there are differences in the immobilised RNA levels between different mRNA targets. GcvB-Cy3 FI was also detected for each of the RNA spots indicating that GcvB-Cy3 interacts with each of the mRNA targets. This was expected based on the results obtained for GcvB probing of the low-density mRNA target array; although the significant binding seen for *ilvI, ilvC* and *hisG,* and weaker binding for the remining targets was not so pronounced in the case for the high-density array results. However, this may simply reflect the level of RNA present on the array. This highlights the need for quantification to normalise the binding seen to the RNA level immobilised so that interaction efficiencies between GcvB and its mRNA targets can be compared.

Normalising the mean FI for the SA-linker bound to the mRNA target spots to the mean FI for the DNA IVT template (Figure 4.23, A) to generate a transcription efficiency confirmed differences between the mRNA targets. This trend was similar to that observed previously. The mean FI for GcvB-Cy3 bound to the mRNA target spots was normalised to the mean FI for the SA-linker bound to the same mRNA target spots to give an interaction efficiency (Figure 4.23, B). This analysis confirmed the qualitative observation that the level of GcvB-Cy3 binding may be mRNA target specific. The highest interaction efficiency was seen for GcvB-Cy3 binding to *ilvC* mRNA target. This was significantly higher than for the other eight mRNA targets.

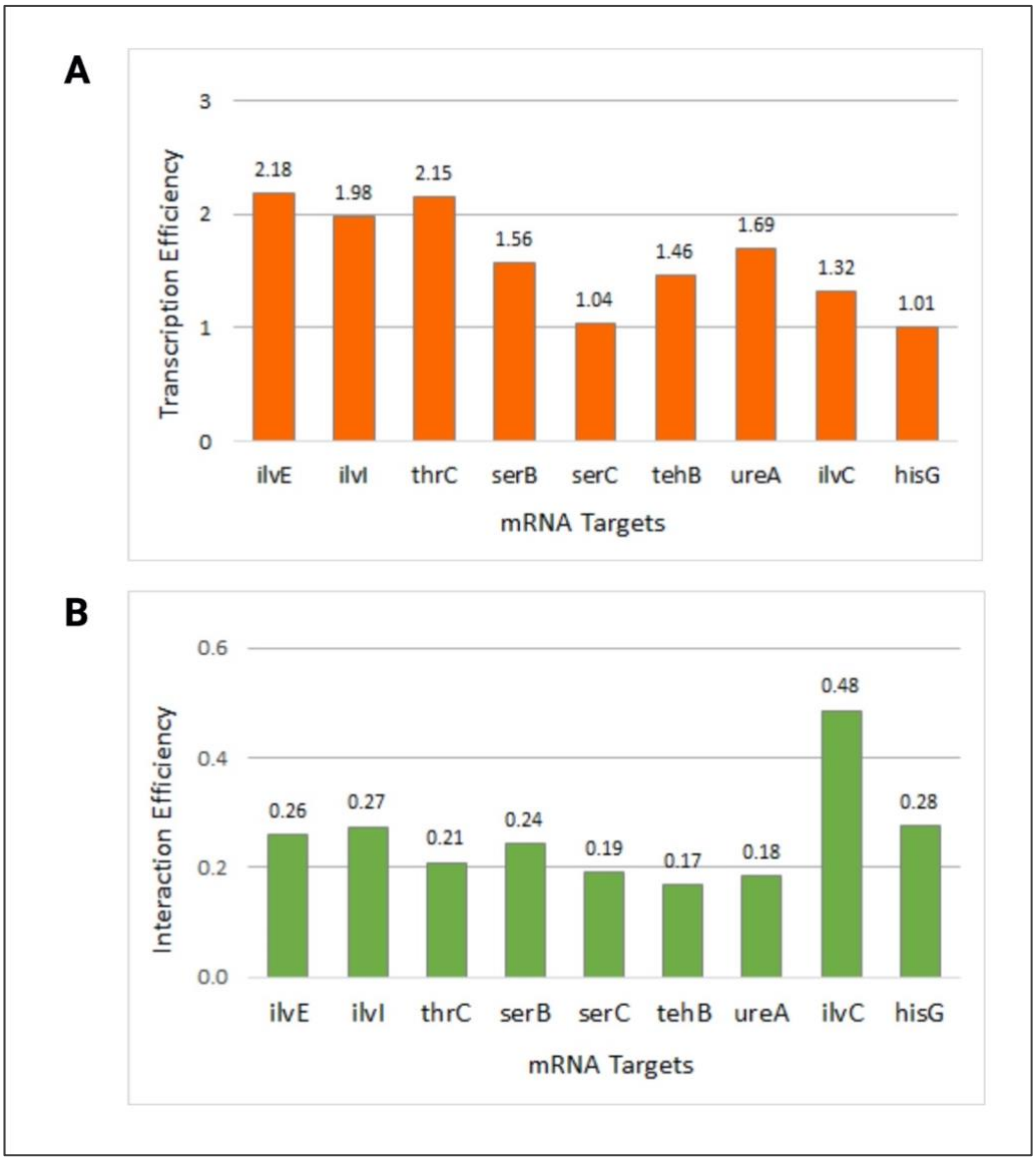


Figure 4.23. Quantification of the probing of a high-density RNA array of putative mRNA targets of GcvB. A) The transcription efficiencies were calculated by normalising the mean FI for the SA-linker bound to the mRNA target spots to the mean FI for the corresponding DNA IVT template spots. B) The GcvB-Cy3-mRNA target interaction efficiencies were calculated by normalising the mean FI for the GcvB-Cy3 bound to the mRNA spot to the mean FI for the SA-linker probe bound to the same spot.

4.5.2. Co-spotting the DNA IVT templates for GcvB and its putative mRNA targets: *in situ* binding of GcvB followed by probing of bound GcvB post-array production with a fluorescently labelled antisense ssDNA probe

A second strategy to investigate the interaction between GcvB and the putative mRNA targets was developed which involved *in situ* binding of GcvB followed by probing of

the bound GcvB, post-array production, with a fluorescently labelled antisense ssDNA probe (GAG-linker probe) (Figure 4.24). The first step of this strategy involved co-spotting DNA IVT templates for both the putative mRNA targets and GcvB (Figure 4.24, 1). Then, during the RNA array generation step (IVT and *in situ* capture), both RNAs are transcribed simultaneously. This allows GcvB to bind to the mRNA target as the RNA is being transcribed, mimicking the *in vivo* scenario more closely than the probing of the mRNA array with GcvB-Cy3 post-array production.

As previously, the mRNA target RNA is conjugated to an SA-aptamer to facilitate capture of the RNA on an SA-coated surface. However, GcvB is not tagged in this manner. This means that GcvB is only immobilised on the RNA array if it binds to the mRNA target (Figure 4.24, 2A). Bound GcvB is then probed post-array production with Alexa488 labelled antisense ssDNA GAG-linker probe and the immobilised mRNA targets are probed post-array production with Alexa647 SA-linker probe (Figure 4.24, 2B). The use of orthogonal fluorophores allows GcvB and the mRNA targets to be detected simultaneously.

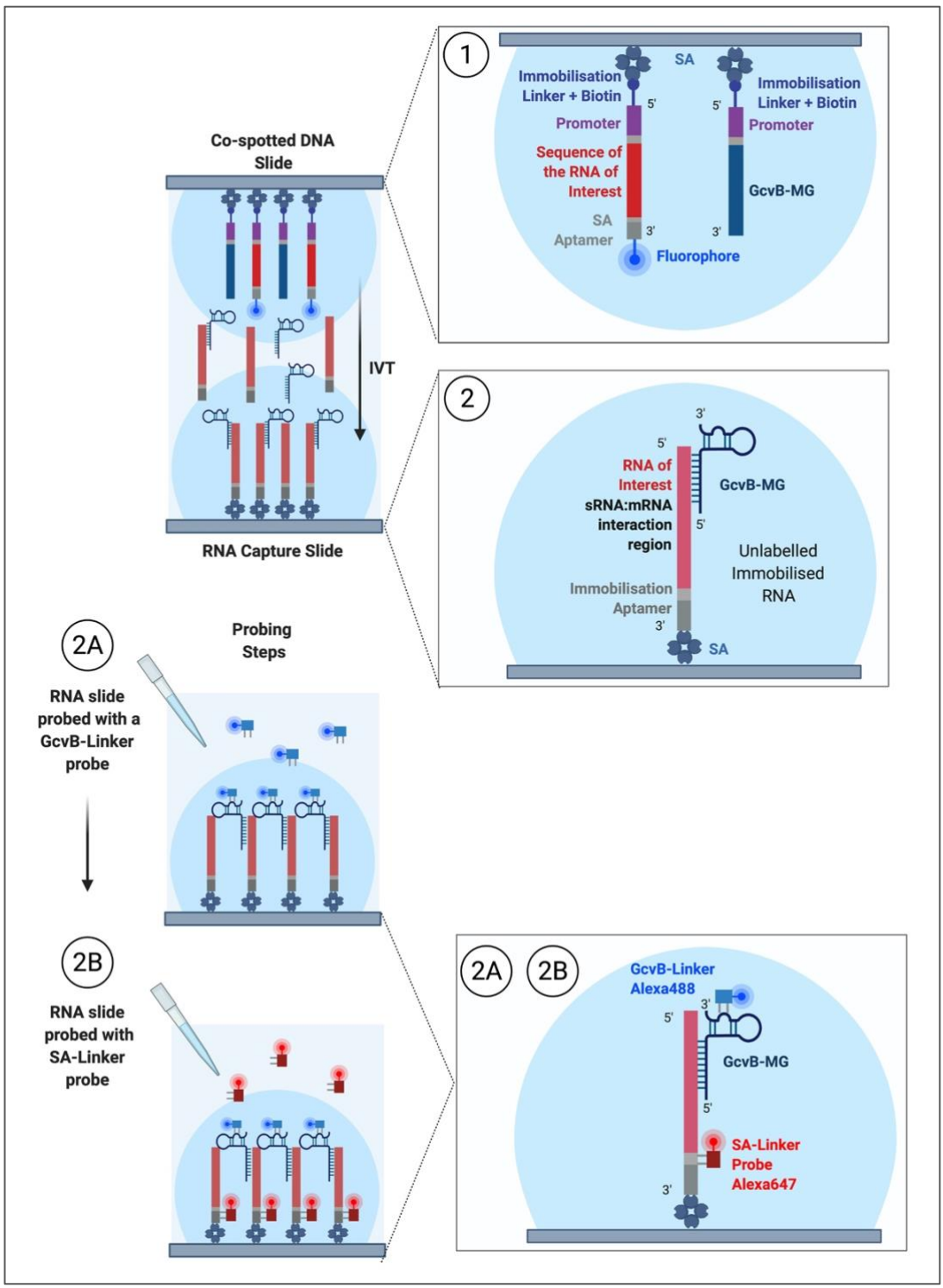


Figure 4.24. Schematic of *in situ* binding of GcvB to mRNA targets followed by probing bound GcvB post-array production with GAG probe. (1) An SA-coated slide is co-spotted with 5′-biotinylated, 3′-Alexa488/Alexa647 labelled DNA IVT templates encoding the putative mRNA targets of GcvB and 5′-biotinylated DNA IVT template encoding GcvB to generate the DNA IVT template array. (2) GcvB and the mRNA targets are transcribed. The mRNA targets are conjugated to an SA aptamer which facilitates immobilisation on the capture slide. GcvB is not conjugated to an SA aptamer and is only immobilised on the capture slide if it binds to the mRNA target. (2A) Bound GcvB is probed post-array production with Alexa488-labelled GAG-linker probe and the mRNA target is probed post-array production with Alexa647/Alexa488-labelled SA-linker probe. The probed array is visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter or using an excitation wavelength of 488 nm and a Standard Blue emission filter and using an excitation wavelength of 635 nm and a Standard Red emission filter.

4.5.2.1. GcvB DNA IVT template design and synthesis

A DNA IVT template was designed for GcvB. The key features and sequence elements of this template are described in Figure 4.25. Similar to the DNA IVT templates for the mRNA targets, the GcvB DNA IVT template consists of a series of functional sequence features that are required for producing the DNA IVT template array, producing the RNA array or visualising the arrays. These include: 1 – A biotinylated linker at the 5′ end to facilitate DNA IVT template immobilisation onto a SA-coated slide; 2 – A T7 bacteriophage RNA polymerase (RNAP) promoter for IVT; 3 – Two guanine nucleotides (GG) to optimise transcription by the T7 RNAP; 4 – Sequence encoding GcvB; 5 – Sequence encoding an unstructured linker to serve as a recognition region for a fluorescently labelled antisense ssDNA probe used for visualisation and quantification of the levels of GcvB bound to the mRNA targets on the RNA array; 6 – A 3′ stem loop to stabilise the transcript. This construct of GcvB was referred to as **M**odified **G**cvB (GcvB-**MG**) from this point onwards, as it includes a specific linker for probing. Key differences between the GcvB-MG DNA IVT template and the DNA IVT templates for the mRNA targets is that there is no SA aptamer to facilitate *in situ* capture of the RNA by an SA-coated capture slide and there is no 3′ fluorophore. This lack of the SA aptamer ensures that GcvB-MG is only immobilised on the RNA array if it is bound to an mRNA target.

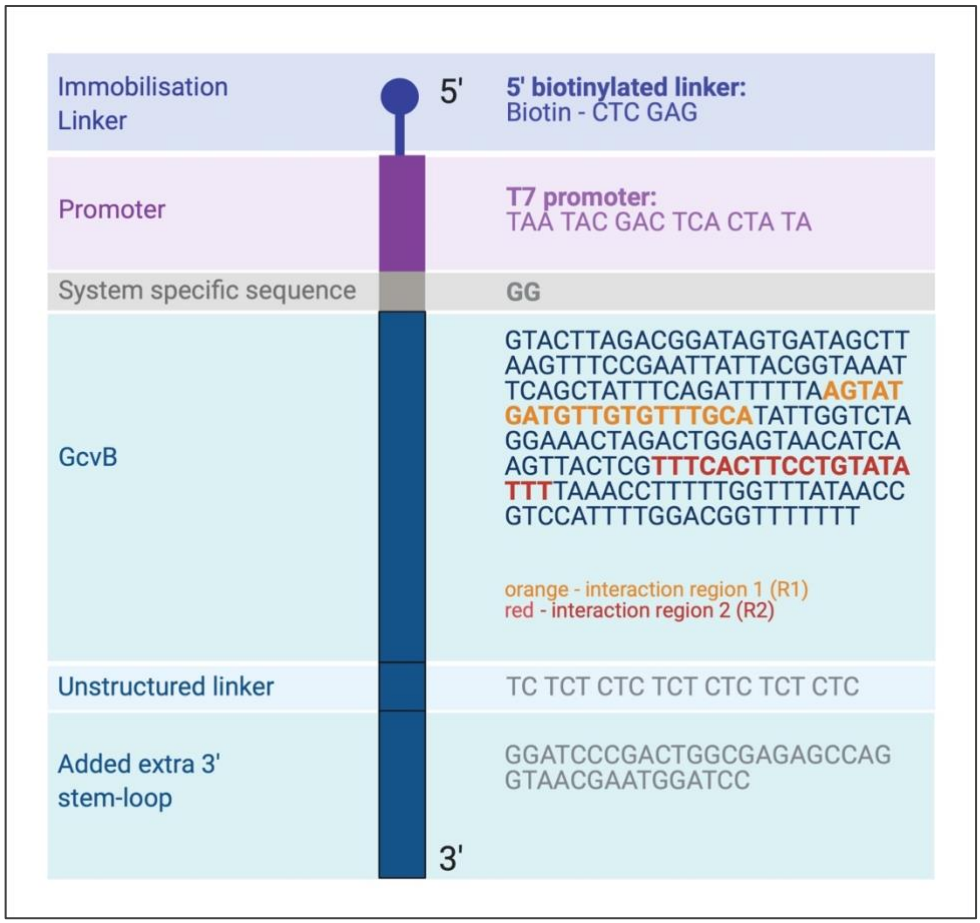


Figure 4.25. Design of the DNA *in vitro* transcription template for GcvB-MG. A schematic of the DNA IVT template. The DNA sequence for each component is indicated. The final 279-nucleotide sequence is: 5′-biotin-ctcgagTAATACGACTCACTATA*GG*GTACTTAGACGGATAGTGATAGCTTAAGTTTCCGAATTATTACGGTAAATTCAGCTATTTCAGATTTTTA**AGTATGATGTTGTGTTTGCA**TATTGGTCTAGGAAACTAGACTGGAGTAACATCAAGTTACTCG**TTTCACTTCCTGTATATTT**TAAACCTTTTTGGTTTATAACCGTCCATTTTGGACGGTTTTTTT*TCTCTCTCTCTCTCTCTCTC***GGATCCCGACTGGCGAGAGCCAGGTAACGAATGGATCC**-3′.

In order to interact with the mRNA targets on the RNA array, GcvB must fold correctly. Therefore, RNAFold was used to predict the secondary structure of GcvB and compare it to the expected RNA product of GcvB-MG. Figure 4.26 shows that the key structural elements predicted to be present in GcvB (SL1-5, R1 and R2) are also predicted to be present in GcvB-MG. It was possible that the addition of an unstructured linker region to GcvB, to create GcvB-MG, could interact with the mRNA targets. Therefore, the IntaRNA webtool was used to determine if interactions would be predicted. The results showed that the additional unstructured linker is unlikely to interact with any of the putative mRNA targets of GcvB (Appendix 8).

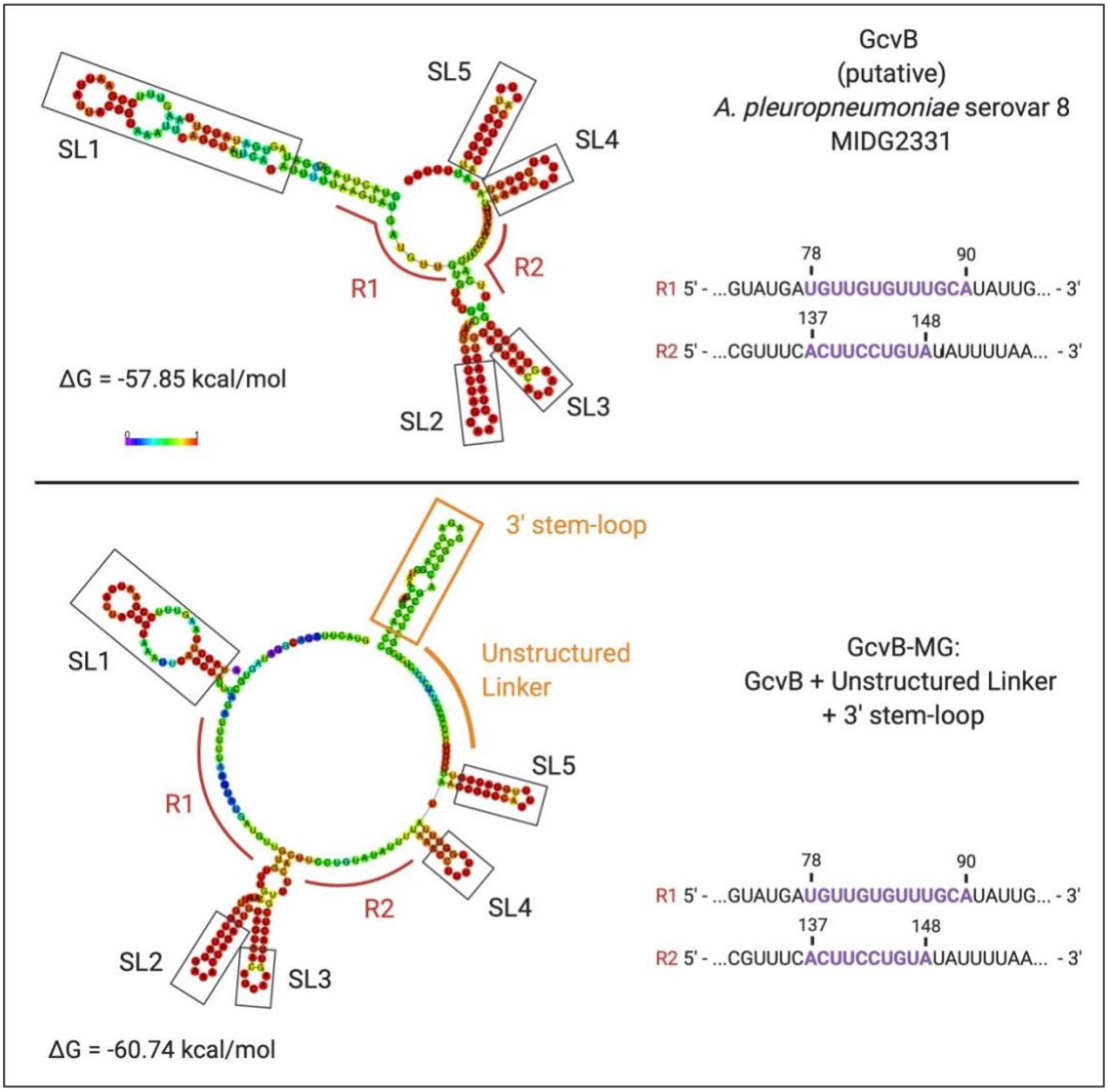


Figure 4.26. Comparison of the predicted secondary structure of GcvB and GcvB-MG. (Top) The predicted secondary structure of GcvB contains 5 stem-loops (grey rectangles labelled SL1-5) and interaction regions in areas with low base-pairing probability (red lines labelled R1 and R2). (Bottom) The predicted secondary structure of GcvB-MG also contains 5 stem-loops (grey rectangles labelled SL1-5) and two interaction regions (red lines labelled R1 and R2). In addition, it contains a 3′ stem-loop (orange rectangle) and an unstructured linker (orange line).

Following DNA IVT template design, gene synthesis primers were designed and obtained from Invitrogen (Thermo Fisher Scientific). The IVT DNA template was synthesised *de novo* in-house by TBIO-PCR using a 5′-biotinylated primer to incorporate the 5′ biotin (Chapter 2, section 2.2.3.2). The size of the DNA IVT template was verified by gel electrophoresis (Appendix 12) and the DNA template produced an IVT product of the expected length in solution (Appendix 13).

4.5.2.2. *In situ* binding of GcvB-MG on low-density arrays followed by probing of bound GcvB-MG post-array production with GAG-linker probe

To confirm that GcvB can bind, and remain bound, to the mRNA targets when it is co-transcribed with its putative mRNA targets, a preliminary experiment was conducted using a low-density RNA array of a subset of the mRNA targets. *ilvI* and *hisG* mRNA targets were selected for this initial experiment because GcvB has consistently interacted with these mRNA targets when added post-array production. A sample (0.2 µl) of 200 nM of the 5′-biotinylated and 3′-Alexa647 labelled *ilvI*, *hisG*, *serC*, *thrC* and *ureA* mRNA target DNA IVT templates and a sample (0.2 µl) of a mixture of 200 nM 5′-biotinylated and 3′-Alexa647 labelled *ilvI* or *hisG* mRNA target DNA IVT template with 200 nM 5′-biotinylated GcvB-MG DNA IVT template were spotted, in duplicate, onto a SA-coated slide, by manual spotting, to generate a DNA IVT template array.

Since each of the mRNA target DNA IVT templates were fluorescently labelled, this allowed visualisation of the DNA IVT template array. The DNA IVT template array was then used to generate an unlabelled mRNA target array. During this step, the co-transcribed GcvB-MG binds to the mRNA targets. The RNA array was probed with 500 nM Alexa488-labelled GAG-linker probe, to detect bound GcvB-MG. Finally, the mRNA target array was probed with 500 nM Alexa488-labelled SA-linker probe, to detect the mRNA targets. A schematic of the experimental design is shown in Figure 4.27.

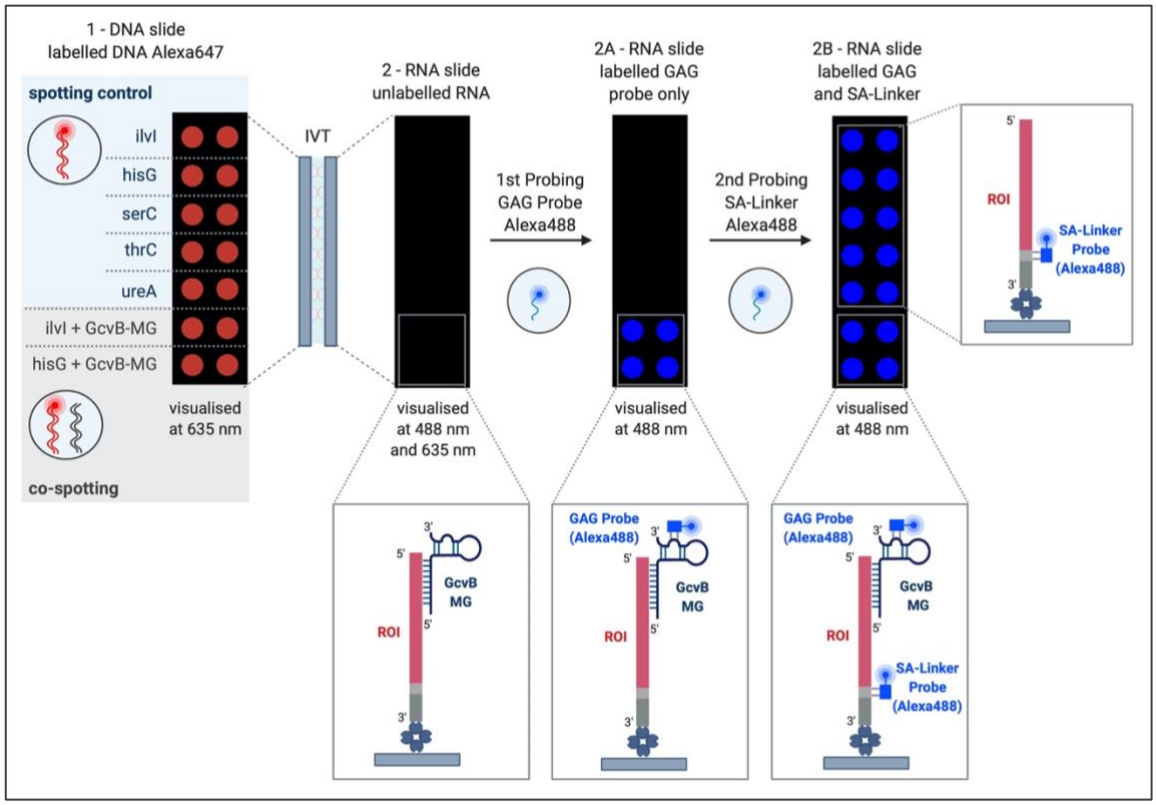


Figure 4.27. Schematic of *in situ* binding of GcvB-MG on low-density arrays followed by probing of bound GcvB-MG post-array production with GAG-linker probe. (1) Samples (0.2 µl) of 200 nM of the 5′-biotinylated and 3′-Alexa647 labelled *ilvI*, *hisG*, *serC*, *thrC* and *ureA* mRNA target DNA IVT templates are manually spotted, in duplicate, on an SA-coated slide, by manual spotting, in the spotting control area. Samples (0.2 µl) of a mixture of 200 nM 5′-biotinylated and 3′-Alexa647 labelled *ilvI* or *hisG* mRNA target DNA IVT template and 200 nM 5′-biotinylated GcvB-MG DNA IVT template are co-spotted, in duplicate, onto a SA-coated slide, by manual spotting, in the co-spotting area. This generates a DNA IVT template array that can be visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. (2) A corresponding unlabelled mRNA target array is generated from the DNA IVT template array in (1) by IVT and *in situ* capture. Co-transcribed unlabelled GcvB-MG binds to the mRNA targets in the co-spotting area. (2A) Bound GcvB-MG is probed with 500 nM antisense ssDNA 5′-Alexa488-GAGAGAGAGAGAGAGAGAGA-3′ (GAG-linker probe) and visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter. (2B) The RNA array is probed with 500 nM Alexa488-labelled SA-linker probe and visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter.

As shown Figure 4.28, 1, each of the DNA IVT templates for the mRNA targets were successfully immobilised to the SA-coated slide. The presence of GcvB-MG does not appear to have affected the immobilisation of the *ilvI* or *hisG* DNA IVT templates. Since the GcvB-MG DNA IVT template is unlabelled, it is not possible to determine if this template has also been immobilised. The GcvB-MG DNA IVT template was not labelled because an orthogonal fluorophore was not available. Prior to probing, no FI was detected for the unlabelled RNA slide (Figure 4.28, 2). Following probing of bound GcvB-MG with GAG-linker probe, significant FI was detected for the *ilvI* and *hisG* mRNA targets in the co-spotting area (Figure

4.28, 2A), indicating that GcvB-MG had been transcribed and bound to the mRNA targets in this area. As expected, no FI was detected for any of the mRNA targets in the control spotting area, presumably because no GcvB-MG was transcribed or bound in this area. To confirm that these mRNA targets were present on the array, and that the lack of GAG-linker probe FI signal in this area is due to a lack of GcvB-MG, the RNA array was probed with Alexa488-labelled SA-linker probe (Figure 4.28, 2B).

Following probing with SA-linker probe, FI signal was present for each of the RNA spots indicating that the mRNA targets were transcribed and captured in the control spotting area, but GcvB-MG was not present. The FI RNA spots in the co-spotting area also appear to increase following probing with the SA-linker probe, presumably due to the SA-linker probe binding to the mRNA targets in this region as well. Overall, this experiment showed that GcvB-MG binds to the *ilvI* and *hisG* mRNA targets when they are co-transcribed.

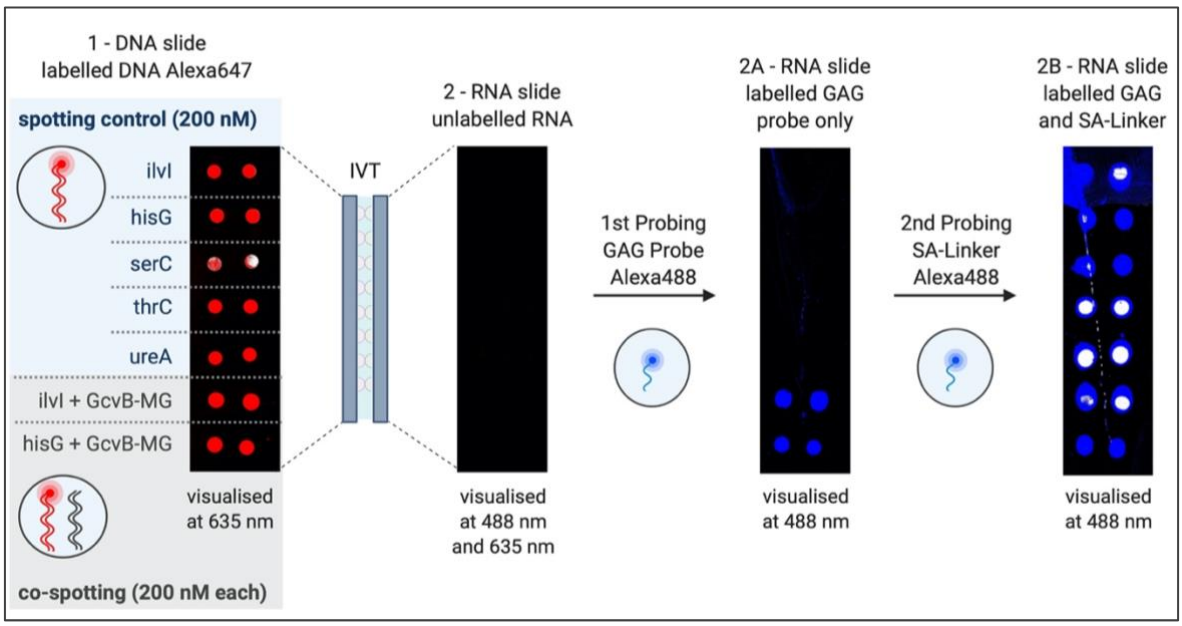


Figure 4.28. *In situ* binding of GcvB-MG on low-density arrays followed by probing of bound GcvB-MG post-array production with GAG-linker probe. (1) Samples (0.2 µl) of 200 nM of the 5′-biotinylated and 3′-Alexa647 labelled *ilvI*, *hisG*, *serC*, *thrC* and *ureA* mRNA target DNA IVT templates were manually spotted, in duplicate, on an SA-coated slide, by manual spotting, in the spotting control area. Samples (0.2 µl) of a mixture of 200 nM 5′-biotinylated and 3′-Alexa647 labelled *ilvI* or *hisG* mRNA target DNA IVT template and 200 nM 5′-biotinylated GcvB-MG DNA IVT template were co-spotted, in duplicate, onto a SA-coated slide, by manual spotting, in the co-spotting area. This generates a DNA IVT template array that can be visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. (2) A corresponding unlabelled mRNA target array is generated from the DNA IVT template array in (1) by IVT and *in situ* capture. Co-transcribed unlabelled GcvB-MG binds to the mRNA targets in the co-spotting area. (2A) Bound GcvB-MG is probed with 500 nM Alexa488-labelled GAG-linker probe and visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter. (2B) The RNA array is probed with 500 nM Alexa488-labelled SA-linker probe and visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter.

4.5.2.3. *In situ* binding of GcvB-MG on high-density arrays followed by probing of bound GcvB-MG post-array production with GAG-linker probe

Having confirmed that GcvB-MG binds to the *ilvI* and *hisG* mRNA targets when they are co-transcribed, the next step was to screen each of the putative mRNA targets of GcvB using this strategy. A high-density array with the layout shown in Figure 4.29 was used for this experiment. The 5′-biotinylated and 3′-Alexa488 labelled DNA IVT template for each of the mRNA targets was spotted, in triplicate, on an SA-coated slide, and at a fixed concentration of 100 nM using an automated arrayer robot. A mixture of 100 nM 5′-biotinylated and 3′-Alexa488 labelled DNA IVT template for each of the mRNA targets together with 100 nM 5′-biotinylated GcvB-MG DNA IVT template was also spotted, in triplicate, for each of the mRNA targets.

Finally, four negative controls were also spotted: 1 – buffer; 2 – 100 nM 5′-biotinylated GcvB-MG DNA IVT template; 3 – 100 nM 5′-biotinylated and 3′-Alexa488 labelled DNA IVT template for a control *ilvC* mRNA target (*ilvC*-S) in which the GcvB-interacting region has been scrambled on the mRNA (Figure 4.30); 4 – a mixture of 100 nM 5′-biotinylated and 3′-Alexa488 labelled DNA IVT template for *ilvC* S together with 100 nM 5′-biotinylated GcvB-MG DNA IVT template.

Since fluorescently labelled DNA IVT templates for the mRNA targets were used, the DNA IVT template array could be visualised. The DNA IVT template array was then used to generate an unlabelled mRNA target array. During this step, co-transcribed GcvB-MG binds to the mRNA targets. The RNA array was probed with 500 nM Alexa488-labelled GAG-linker probe, to detect bound GcvB-MG. Finally, the mRNA target array was probed with 500 nM Alexa647-labelled SA-linker probe, to detect the mRNA targets. A schematic of the experimental design is shown in Figure 4.29.

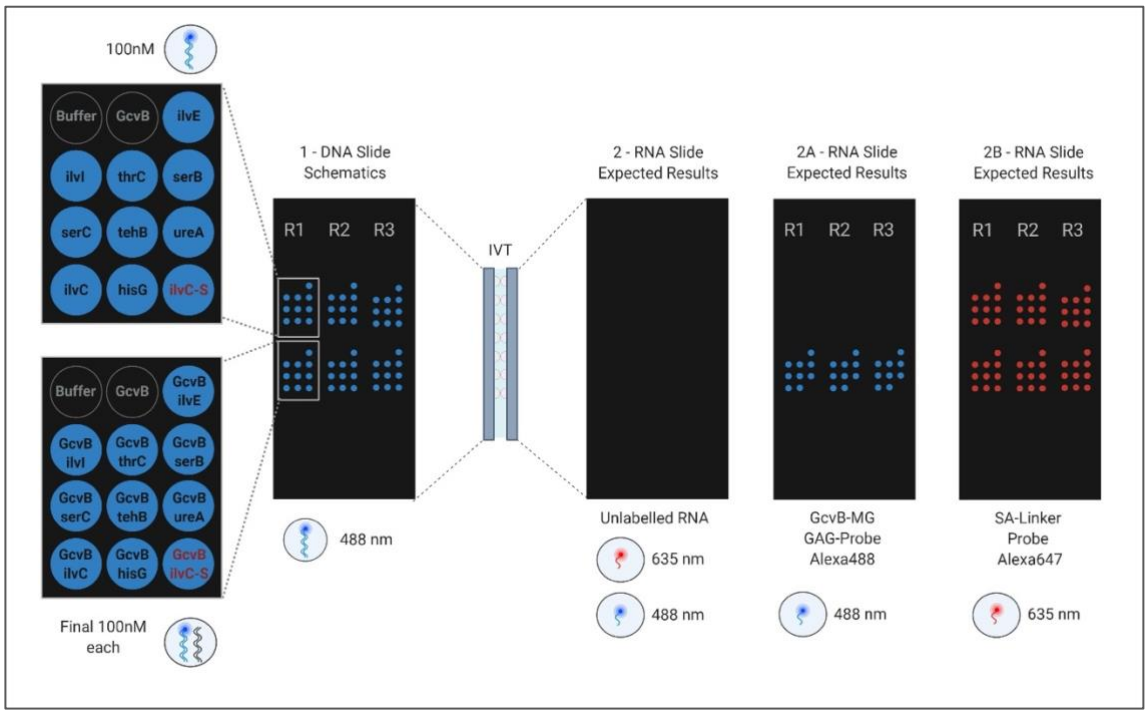


Figure 4.29. Schematic of *in situ* binding of GcvB on high-density arrays followed by probing of bound GcvB post-array production with GAG-linker probe. (1) (Top) Three fields (R1, R2 and R3) of 3 x 4 spots were spotted with 100 nM 5′-biotinylated and 3′-Alexa488 labelled mRNA target DNA IVT template with an automated arrayer robot. (Bottom) Three fields (R1, R2 and R3) of 3 x 4 spots were spotted with a mixture of 100 nM 5′-biotinylated and 3′-Alexa488 labelled mRNA target DNA IVT template together with 100 nM 5′-biotinylated GcvB-MG DNA IVT template with an automated arrayer robot. A buffer control and a 100 nM 5′-biotinylated GcvB-MG DNA IVT template control were spotted in all six fields. The DNA IVT template array was visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter. (2) A corresponding unlabelled mRNA target array is generated from the DNA IVT template array in (1) by IVT and *in situ* capture. (2A) GcvB-MG bound to the mRNA target array is probed with 500 nM Alexa488-labelled GAG-linker probe and visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter. (2B) The RNA array is probed with 500 nM Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter.

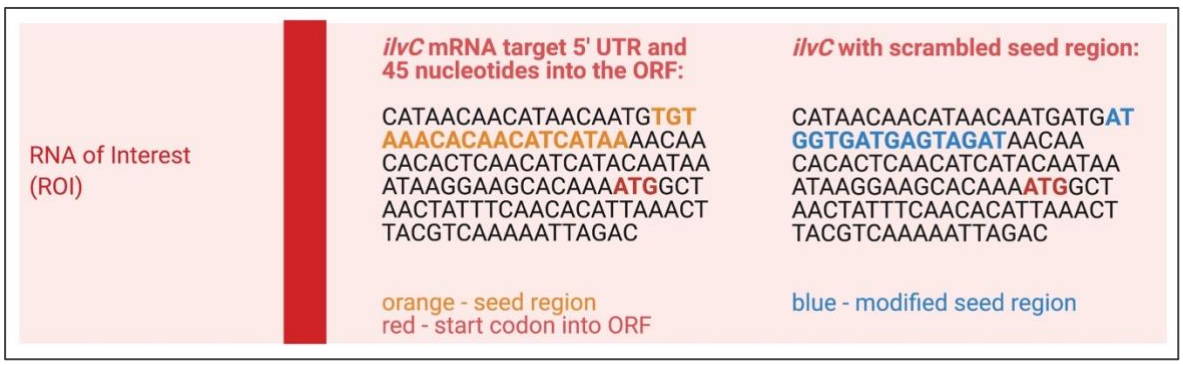


Figure 4.30. Design of the DNA *in vitro* transcription template for the *ilvC*-S mRNA. The DNA IVT template for the *ilvC*-S mRNA has the same general architecture as that presented in Figure 4.2. The RNA of interest (ROI) region for this DNA IVT template is shown with the sequence for *ilvC* (left) and *ilvC*-S (right). The seed region of *ilvC* that is predicted to interact with GcvB (orange), as predicted in TargetRNA2, has been scrambled in *ilvC*-S (blue).

As shown Figure 4.31, 1, visualisation of the DNA IVT template array reveals FI for each of the spots containing immobilised 5′-biotinylated and 3′-Alexa488 labelled mRNA target DNA IVT template, including the *ilvC*-S mRNA. As expected, no FI was detected for the buffer spots or the spots containing immobilised unlabelled 5′-biotinylated GcvB-MG DNA IVT template. The presence of GcvB-MG does not appear to have adversely affected the immobilisation of the mRNA target DNA IVT templates. Since the GcvB-MG DNA IVT template is unlabelled, it is not possible to determine if this template has also been immobilised. Prior to probing, no FI was detected for the unlabelled RNA slide generated from the DNA IVT template array by IVT and RNA capture (Figure 4.31, 2). Following probing of bound GcvB-MG with Alexa488-labelled GAG-linker probe, significant FI was detected for each of the RNA spots generated by co-transcription of the mRNA target and GcvB-MG, except for the RNA spot generated by co-transcription of *ilvC*-S mRNA and GcvB-MG (Figure 4.31, 2A). This indicates that GcvB-MG had been transcribed and has bound to each of the mRNA targets except for *ilvC*-S mRNA.

The level of bound GcvB detected was dependent on the mRNA target. Also, as expected, no FI was detected for the buffer spots and the GcvB-MG spots, suggesting that the GcvB-MG RNA is not captured directly by the SA-coated slide, or any of the mRNA targets transcribed in the absence of GcvB-MG. To confirm that the mRNA targets transcribed in the absence of GcvB-MG, and *ilvC*-S mRNA transcribed in the presence of GcvB-MG are immobilised on the RNA array, the array was probed with Alexa647-labelled SA-linker probe (Figure 4.31, 2B). The use of orthogonal fluorophores for the GAG-linker probe and the SA-linker probe allowed independent detection of the mRNA target co-transcribed with GcvB-MG.

Following probing with SA-linker probe, FI signal was present for each of the RNA spots, except for the buffer spots and the GcvB-MG spots. This indicates that each of the mRNA targets that were expected to be transcribed and captured were transcribed and captured. The levels of RNA that were transcribed and captured were dependent on the mRNA target, as observed previously. Co-transcription of GcvB-MG does not appear to have significantly affected this trend. Overall, this experiment showed that GcvB-MG binds to each of the predicted mRNA targets when they are co-transcribed.

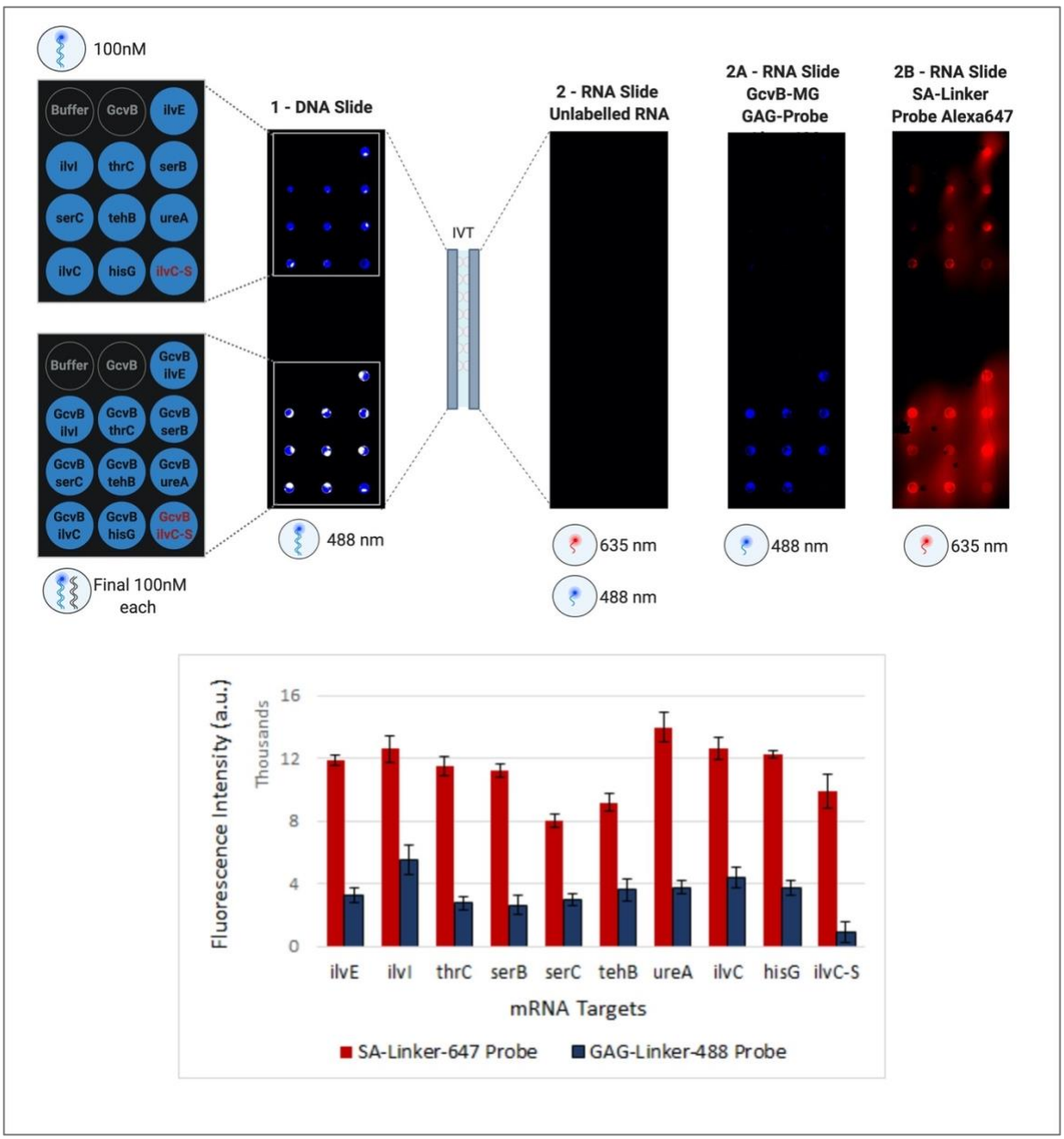


Figure 4.31 *In situ* binding of GcvB on a high-density array followed by probing of bound GcvB post-array production with GAG-linker probe. (1) (Top) A representative field of 3 x 4 spots spotted with 100 nM 5′-biotinylated and 3′-Alexa488 labelled mRNA target DNA IVT template with an automated arrayer robot. (Bottom) A representative field of 3 x 4 spots spotted with a mixture of 100 nM 5′-biotinylated and 3′-Alexa488 labelled mRNA target DNA IVT template and 100 nM 5′-biotinylated GcvB-MG DNA IVT template with an automated arrayer robot. A buffer control and a 100 nM 5′-biotinylated GcvB-MG DNA IVT template control were spotted in both fields. The DNA IVT template array was visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter. (2) The corresponding fields of the unlabelled mRNA target array generated from the DNA IVT template array in (1) by IVT and *in situ* capture. (2A) GcvB-MG bound to the mRNA target array probed with 500 nM Alexa488-labelled GAG-linker probe and visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter. (2B) The RNA array probed with 500 nM Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. A plot of mean FI (in a.u.) for SA-linker probe bound to the mRNA target (red bars) and for GAG-linker probe bound to GcvB-MG bound to the same mRNA target (blue bars). Data are the mean of 3 spots.

Normalising the mean FI for the SA-linker bound to the mRNA target spots to the mean FI for the DNA IVT template (Figure 4.32, A) confirms that the levels of the immobilised mRNA targets differ. The mean FI for GAG-linker probe, bound to GcvB-MG, bound to the mRNA target, was normalised to the mean FI for the SA-linker bound to the same mRNA target to give an interaction efficiency (Figure 4.32, B). This analysis confirmed the observation that the level of GcvB-MG binding may be mRNA target specific. The trend in interaction efficiencies for GcvB-MG binding was similar to the trend observed for GcvB-Cy3 binding. Therefore, similar results were obtained whether GcvB bound to the mRNA target post-array production or *in situ* during co-transcription and mRNA target capture.

The highest interaction efficiency for GcvB-MG binding was to the mRNA target *ilvI*, closely followed by *ilvC.* GcvB-MG interaction efficiencies for the other seven mRNA targets were seen to be lower. As expected, the interaction efficiency for GcvB-MG binding to the control mRNA, *ilvC*-S was significantly reduced compared to the test mRNAs, but not zero. The relative ratio between *ilvC* and *ilvC*-S indicated a 9.5-fold decrease. Despite IntaRNA not predicting an interaction between GcvB seed region and *ilvC-S* scrambled interaction region, visual inspection of the *ilvC*-S sequence shows it contains three CACAA pentanucleotide repeats which are complementary to part of the GcvB seed region:

(CATAACAACATAACAATG**ATGATGGTGATGAGTAGAT**AACAACACACTCAACATCATACAATAA ATAAGGAAGCACAAAATGGCTAACTATTTCAACACATTAAACTTACGTCAAAAATTAGAC), which may account for the minimal binding seen.

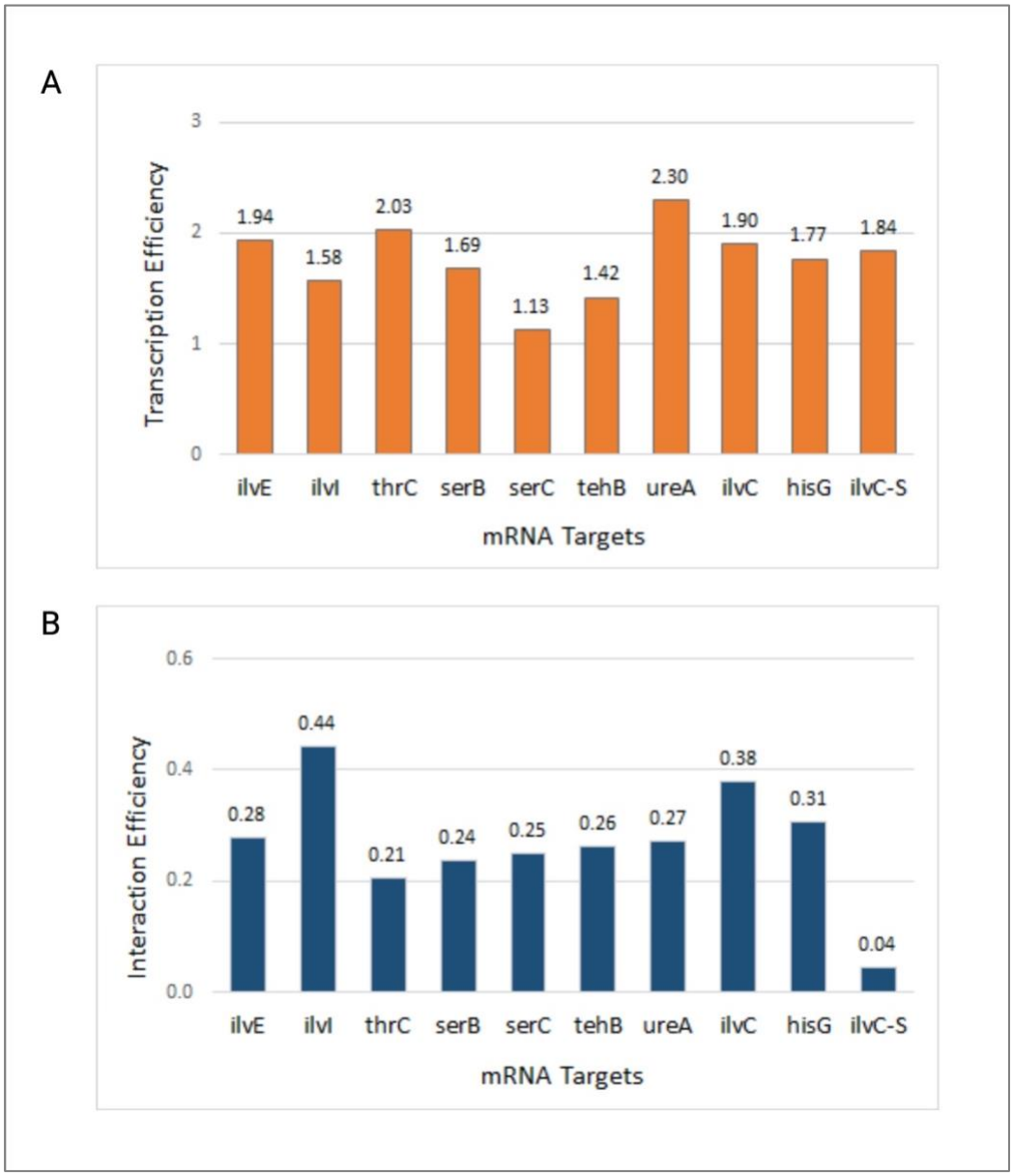


Figure 4.32. Quantification of *in situ* binding of GcvB on a high-density array by probing of bound GcvB post-array production with GAG-linker probe. (A) The transcription efficiencies were calculated by normalising the mean FI for the SA-linker bound to the mRNA target spots to the mean FI for the corresponding DNA IVT template spots. (B) The GcvB-MG interaction efficiency was calculated by normalising the mean FI for the GAG-linker probe, bound to GcvB-MG, bound to the mRNA target, to the mean FI for the SA-linker probe bound to the same mRNA target.

To further investigate the relative interaction efficiency of GcvB-MG binding to each of the predicted mRNA targets preliminary experiments were conducted that involved co-spotting a fixed concentration of DNA IVT template for the mRNA target and a concentration range of GcvB-MG. Four mRNA targets were selected to test this approach, *ilvI, ilvC, hisG* and *ilvC*-S. *ilvI, ilvC* and *hisG* were chosen because GcvB-MG bound to these mRNA targets with the highest interaction efficiencies in previous experiments, and *ilvC*-S

provides a control. A mixture of 100 nM 5′-biotinylated and 3′-Alexa647 labelled DNA IVT template for the *ilvI*, *ilvC*, *hisG* and *ilvC*-S mRNA targets and 0, 5, 10, 25, 50, 100, 125, 150, 200, 250, 250 or 300 nM 5′-biotinylated GcvB-MG DNA IVT template was spotted, in triplicate, on an SA-coated slide, using an automated arrayer robot, to generate an DNA IVT template array. This was visualised using the Alexa647 fluorophore. The layout of the DNA IVT template array is shown in Figure 4.33.

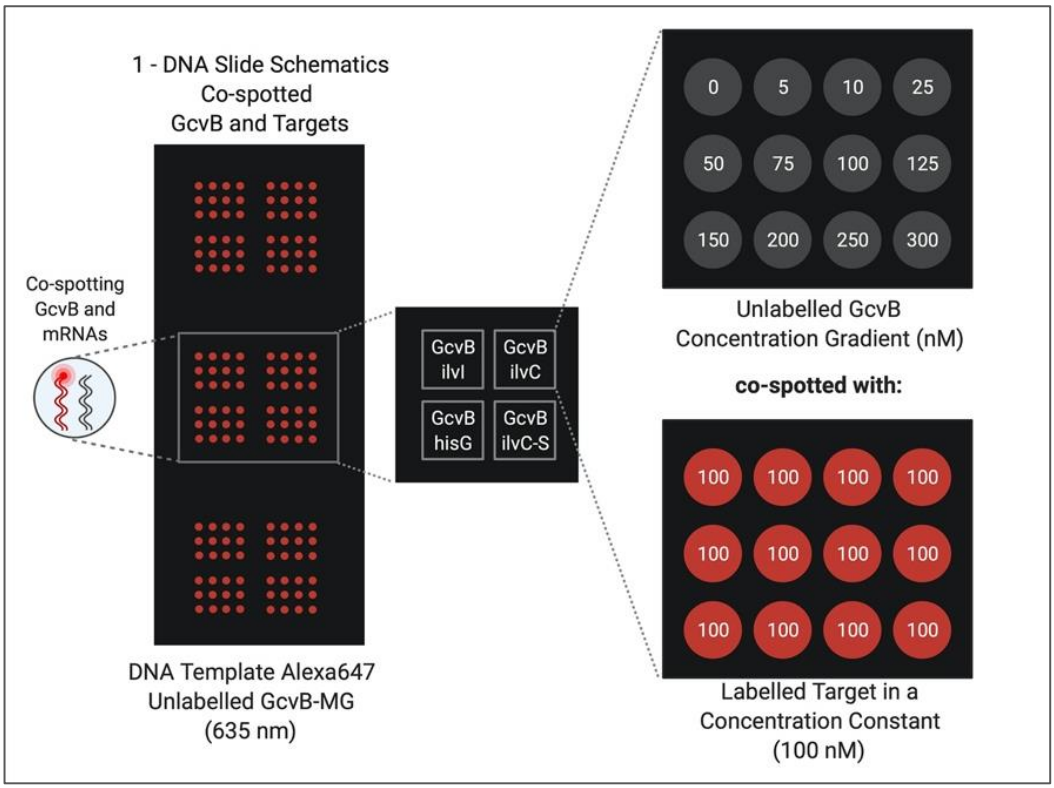


Figure 4.33. Schematic of the layout of the high-density DNA IVT template array for *in situ* binding of a concentration range of GcvB. (1) Three grids of 2 x 2 fields of 4 x 3 spots were spotted with a mixture of 100 nM 5′-biotinylated and 3′-Alexa647 labelled *ilvI*, *ilvC*, *hisG* or *ilvC*-S mRNA DNA IVT template and 0, 5, 10, 25, 50, 100, 125, 150, 200, 250, 250 or 300 nM 5′-biotinylated GcvB-MG DNA IVT template with an automated arrayer robot. The DNA IVT template array is visualised using an excitation wavelength of 647 nm and a Standard Red emission filter.

This DNA IVT template array was used to generate an unlabelled RNA array by IVT and RNA capture on an SA-coated slide. During this step, co-transcribed GcvB-MG binds to the mRNA targets. The RNA array was then probed with 500 nM Alexa488-labelled GAG-linker probe, to detect bound GcvB-MG and with 500 nM Alexa647-labelled SA-linker probe, to detect the mRNA targets. As shown in Figure 4.34, A, visualisation of the DNA IVT template array reveals FI for each of the spots containing immobilised 5′-biotinylated and 3′-Alexa647 labelled mRNA target DNA IVT template. Following probing of bound GcvB-MG

with Alexa488-labelled GAG-linker probe, increasing FI was detected for the RNA spots generated by co-transcription of the *ilvI*, *ilvC* and *hisG* mRNA targets and GcvB-MG with increasing concentration of spotted GcvB-MG DNA IVT template (Figure 4.34, 2A). No FI was visible for the RNA spots generated by co-transcription of *ilvC*-S mRNA and GcvB-MG. For the *ilvI* mRNA target, FI was detected for 5 nM GcvB-MG DNA IVT template (the lowest concentration spotted) and above. For the *ilvC* and *hisG* mRNA targets, FI was not detected until approximately 50 nM GcvB-MG DNA IVT template. This suggests that the efficiency of GcvB-MG binding may depend on the mRNA target. To confirm that the differential binding was not due to the mRNA target level on the RNA array, the array was probed with Alexa647-labelled SA-linker probe (Figure 4.34, 2B). Following probing with SA-linker probe, FI signal was present for each of the RNA spots.

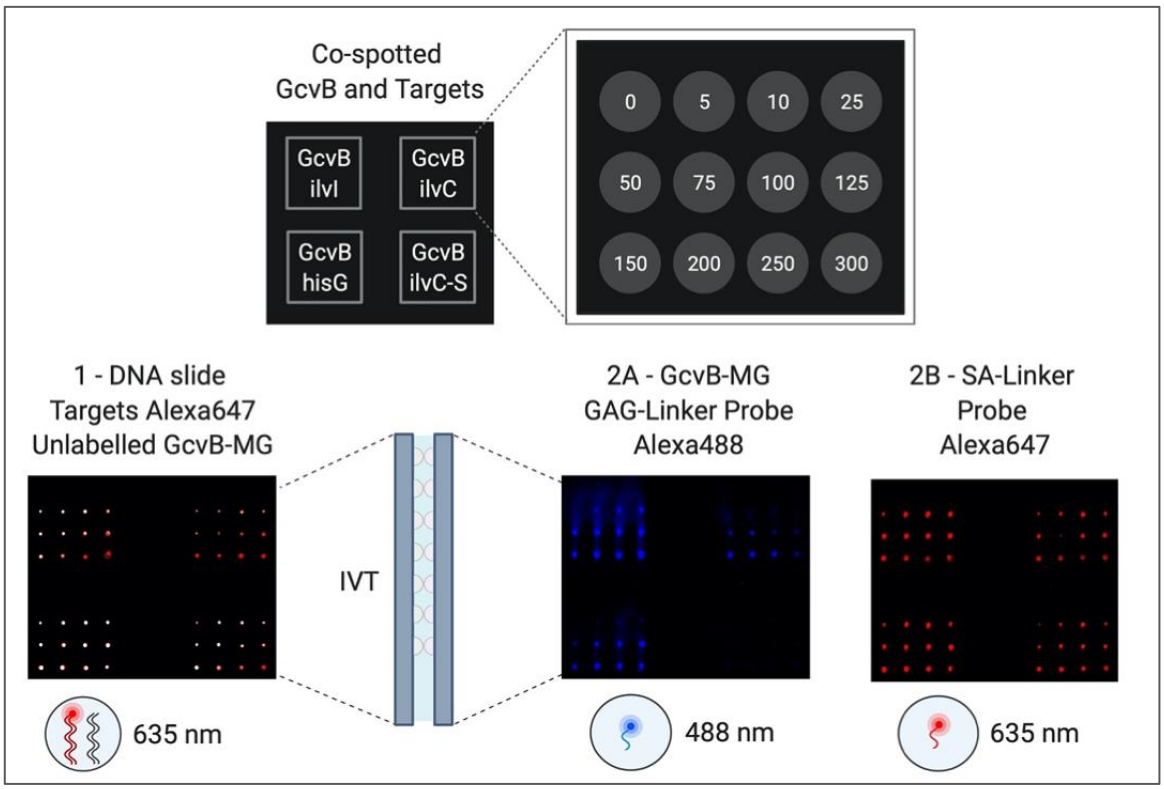


Figure 4.34. *In situ* binding of a concentration range of GcvB to mRNA targets. (1) A representative grid of 2 x 2 fields of 4 x 3 spots spotted with a mixture of 100 nM 5′-biotinylated and 3′-Alexa647 labelled *ilvI*, *ilvC*, *hisG* or *ilvC*-S mRNA DNA IVT template and 0, 5, 10, 25, 50, 100, 125, 150, 200, 250, 250 or 300 nM 5′-biotinylated GcvB-MG DNA IVT template with an automated arrayer robot. The DNA IVT template array was visualised using an excitation wavelength of 635 nm and a Standard Red emission filter. (2A) GcvB-MG bound to the corresponding mRNA target array probed with 500 nM Alexa488-labelled GAG-linker probe and visualised using an excitation wavelength of 488 nm and a Standard Blue emission filter. (2B) The RNA array probed with 500 nM Alexa647-labelled SA-linker probe and visualised using an excitation wavelength of 635 nm and a Standard Red emission filter.

To investigate the relative binding of GcvB-MG to *ilvI*, *ilvC* and *hisG* further, the mean Alexa488 GAG-linker probe FI, representing the level of bound GcvB-MG, was normalised to the mean Alexa647 SA-linker probe FI, representing the mRNA target level, and plotted against the concentration of GcvB-MG DNA IVT template that was spotted (Figure 4.35). The level of GcvB-MG bound appears to increase with increasing concentration of GcvB-MG DNA IVT template, before reaching a plateau. The level of the plateau is highest for *ilvI*, followed by *ilvC* and lowest for *hisG.*

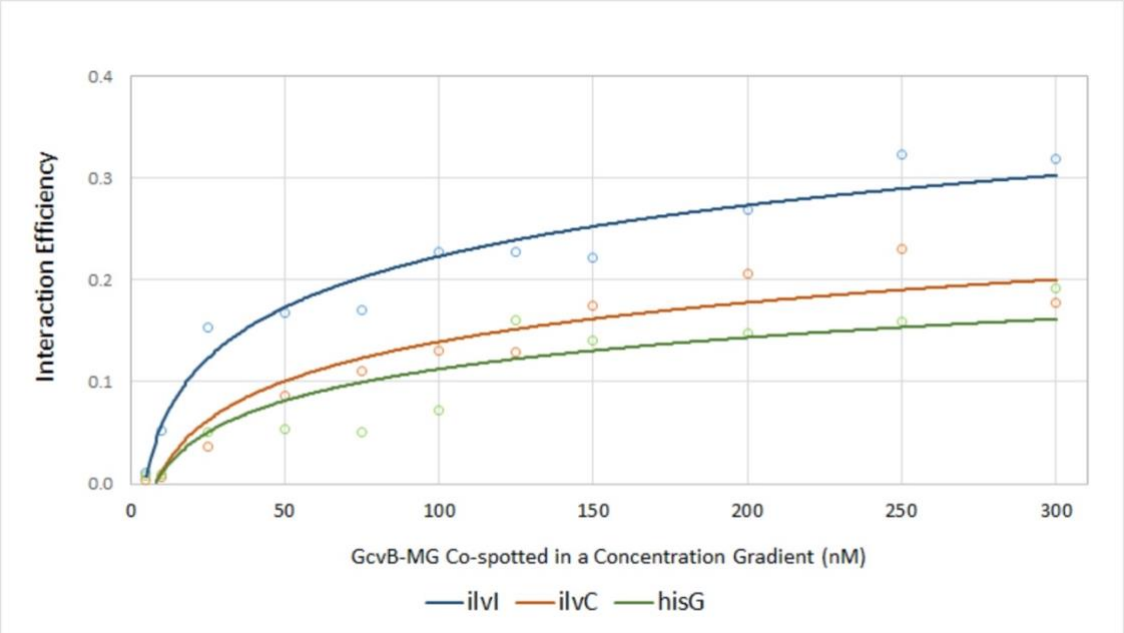


Figure 4.35. Analysis of *in situ* binding of a concentration range of GcvB to mRNA targets. Mean Alexa488 GAG-linker probe FI, normalised to the mean Alexa647 SA-linker probe FI, plotted against the concentration of GcvB-MG DNA IVT template that was spotted. Logarithmic trendlines have been added as a visual aid only.

4.6. Summary and conclusions

In this chapter, the novel RNA array platform that has been developed in the Callaghan group at the University of Portsmouth, has been applied to the validation of the GcvB-mRNA target interactions that were predicted in Chapter 3. Initially, an RNA array of the nine putative mRNA targets of GcvB was designed and synthesised. Analysis of the RNA levels on this RNA array indicated that each of the RNAs may be transcribed/captured with a different transcription efficiency. This suggests that the observed transcription efficiency is a property of the DNA IVT template/mRNA, or RNA capture efficiency by the transcribed RNA conjugate. This variability in the RNA levels for different RNA species should therefore be considered when analysing interaction data.

Having successfully generated an RNA of putative mRNA targets of GcvB, two strategies were devised to screen them for GcvB binding. The first strategy involved probing the RNA array post-array production with *in vitro* transcribed GcvB. This strategy has been used successfully for other sRNA-mRNA interactions and RNA-small molecule interactions (Phillips *et al*., 2018; Henderson *et al*., 2019). Using this strategy, GcvB bound to each of the putative mRNA targets identified in Chapter 3. Once the levels of mRNA target RNA had been accounted for, differences in interaction efficiencies were clearly identified for GcvB binding to each of the mRNA targets.

The second strategy to investigate GcvB binding was to co-transcribe the GcvB with the mRNA target to allow binding of GcvB to the mRNA target *in situ.* This strategy may more closely mimic the situation *in vivo* and has also been used successfully to investigate regulatory RNA interactions (Norouzi *et al*., 2019). This strategy also suggested interaction between GcvB and each of the predicted mRNA targets. Further, the interactions identified appears to be specific as a control mRNA target, in which the GcvB-interacting region had been scrambled, bound GcvB with a significant decreased rate.

Similar to the results obtained when adding GcvB post-array production, the data suggested that there are GcvB-interaction preferences between the mRNA targets which warrant further study. As discussed in the Introduction Chapter, due to their role in bacterial pathogenesis, sRNAs potentially represent a novel therapeutic target. Having experimentally validated that GcvB interacts with the predicted mRNA targets, with recognised links to bacterial virulence, the following Chapter will explore targeting the

GcvB with an antisense nucleic acid mimic (NAM) to disrupt these interactions as a potential therapeutic strategy to address disease caused by APP.

Chapter 5

5. Analysis of the use of a PNA molecule to disrupt sRNA-mRNA interactions

5.1. Introduction

Section 1.3.3 in the Introduction Chapter discussed the possibility of therapeutic targeting of bacterial small RNAs (sRNAs) due to their role in bacterial virulence. Among sRNA types, bacterial *trans*-acting sRNAs regulate gene expression by targeting a set of genes located in distant loci in the genome, which are often involved in similar metabolic pathways that may be linked to virulence. This type of post-transcriptional regulation occurs when the sRNA binds to specific sequences of their target mRNAs, altering translation through post-transcriptional mechanisms. The sRNA-mRNA base-pairing region consists of imperfect complementarity of a short stretch of sequence, known as the seed region.

Following *in silico* analysis in Chapter 3, a seed region has been predicted to be present in a single stranded region of GcvB. Consecutive target prediction analysis of GcvB confirmed that a set of putative mRNAs linked to pathogenicity in APP may interact with GcvB in this predicted seed region. The interaction validation pipeline carried out in Chapter 4 revealed that all of the nine selected predicted mRNA targets interact with the GcvB with differing interaction efficiencies. These outputs lead to the second part of the *in vitro* experiments regarding exploring GcvB-mRNA target interaction disruption. Therefore, the aim of this final chapter was to specifically disrupt GcvB pairing to mRNA targets using a nucleic acid mimic (NAM) which binds specifically to the GcvB seed region (Figure 5.1). Evidence that a nucleic acid mimic can disrupt GcvB pairing to mRNA targets responsible for virulence could suggest a possible ability to inhibit such pathways *in vivo,* as a possible therapeutic strategy.

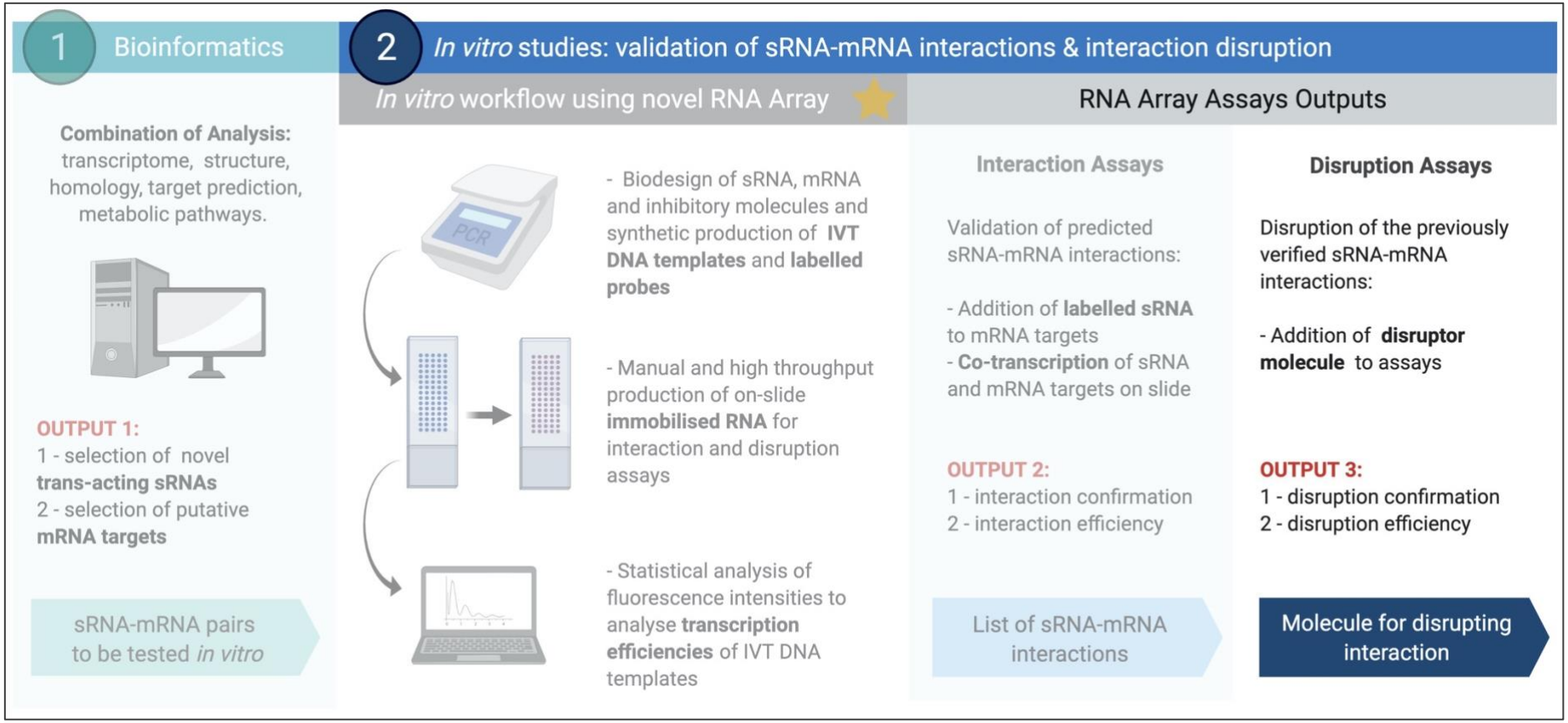


Figure 5.1. Interaction disruption assays. The predicted GcvB-mRNA interactions were validated by the interaction assays. The final part of this study utilises the RNA Array method to explore the disruption of the validated GcvB-mRNA interactions using a final strategy where a specifically designed PNA molecule is added to the established interaction assays. The expected output is an effective PNA molecule that disrupts these interactions.

5.1.1. Antisense inhibition of sRNAs by PNAs

The concept of antisense inhibition of sRNAs is based on the possibility of interfering in the regulation of gene expression in the bacterial cell, by modulating the mechanisms of mRNA regulation by sRNAs. As reviewed in Chapter 1 (section 1.3.3), antisense inhibition technology by NAMs, such as PNA, has been previously used in prokaryote cells to target mRNAs and inhibit their expression. However, the use of antisense NAMs for targeting *trans*-acting sRNAs, instead of mRNAs, is a more versatile strategy, as it enables the regulation of an entire metabolic pathway to be controlled through the disruption of the functionality of a single sRNA. This is because *trans*-acting sRNAs often interact with a multitude of mRNA targets linked to a bacterial pathway, such as those involved in virulence or pathogenicity, thus only one NAM specific to the seed region of the targeted sRNA would be necessary to affect a pathway, replacing the necessity to introduce a NAM for each targeted mRNA.

Antisense targeting of sRNAs by NAMs works by blocking the seed region of the sRNA, preventing its mechanisms of regulation, as shown in Figure 5.2. Different to the imperfect base-pairing of sRNA-mRNA interactions, a NAM may be designed to pair with a given sRNA with perfect complementarity to its full or partial seed region sequence, offering a high affinity interaction which may favour NAM binding, instead of the mRNA, in a competition for the sRNA seed region. Once the NAM specifically binds to the complementary region in the sRNA, it disrupts seed region functionality, which may eventually lead to sRNA degradation by nucleases.

A short literature review on antisense therapy in Chapter 1 (section 1.3.3) introduced a few types of NAMs currently used in biotechnology. However, more commonly used in antibacterial studies are peptide nucleic acids (PNAs), and morpholinos (MOs) (Good *et al*., 2001; Geller *et al*., 2003; reviewed in Summerton, 2006; Verona *et al*., 2017). Like morpholinos, PNAs are uncharged nucleic acid mimics that lack a deoxyribose or ribose sugar backbone. The modified backbone of PNAs and MOs increases protection against enzymatic degradation, making them an ideal option for *in vivo* as well as *in vitro* studies. However, a higher affinity for RNA molecules is observed in PNAs in comparison to MOs (Summerton, 2006). For this study, PNA was chosen as the most suitable option of

antisense molecule due to its chemical simplicity, biological stability, and high affinity hybridisation with RNA.

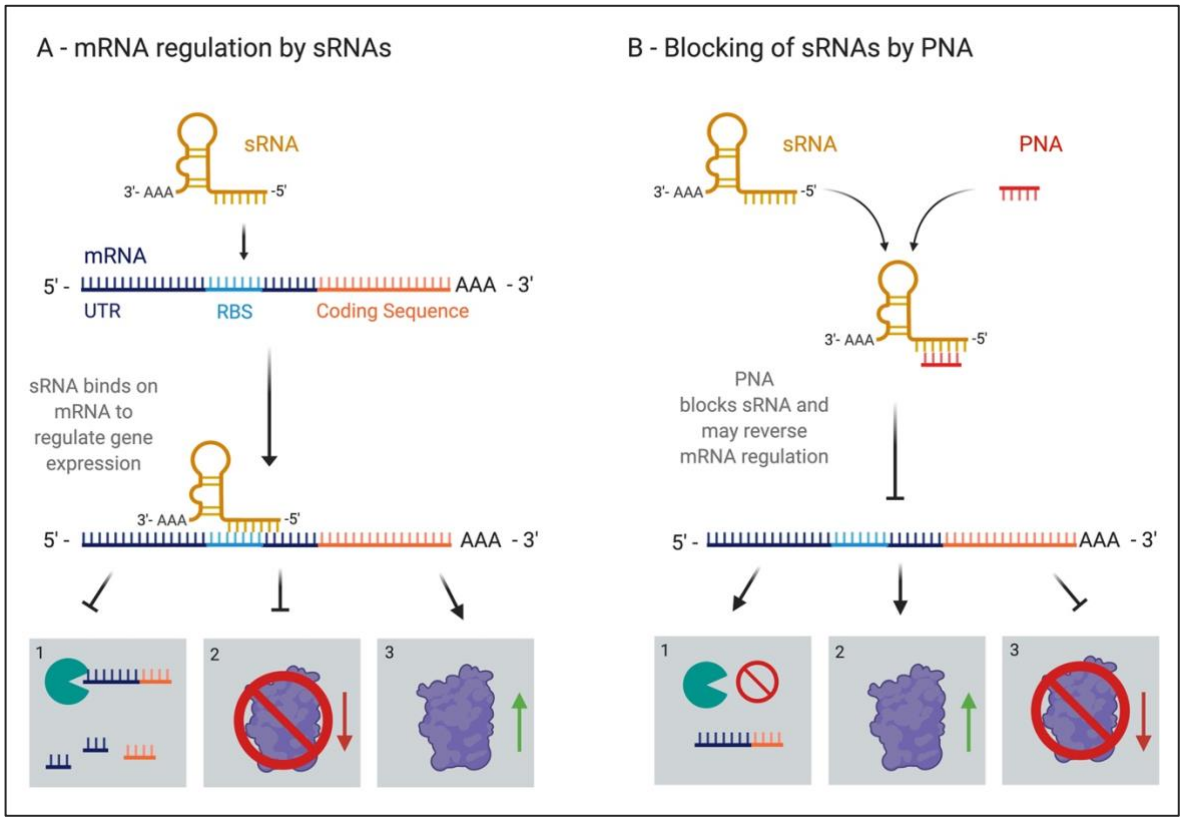


Figure 5.2. Controlling bacterial gene expression by targeting sRNAs with antisense NAMs. A) In bacterial cells, post-transcriptional regulation of mRNAs is often regulated by interaction with the seed region of sRNAs through a variety of mechanisms, resulting in the mRNA-sRNA duplex degradation by nucleases (1), concealing of the ribosomal binding site (RBS) inhibiting mRNA translation (2), or change of the mRNA structural conformation to expose its RBS and facilitate translation (3). These mRNAs may be involved in metabolic pathways linked to virulence, thus the versatile regulation by sRNAs may promote the amplification of bacterial pathogenicity. B) Inhibiting the sRNA seed region by antisense NAMs may revert mRNA regulation in the opposite direction of its original regulation; discontinue the recruiting of nucleases prolonging mRNA availability for translation by ribosomes (1), halt influence on the RBS of the mRNA for up (2) or down (3) regulation of a specific protein.

5.2. Method development to test PNA disruption of GcvB-mRNA interactions

In the previous chapter, separate co-spotting of IVT DNA templates of modified GcvB (GcvB-MG) with each of the nine predicted target mRNAs has been established as the preferred RNA Array method to study RNA-RNA interactions. This is due to the flexibility of the approach in testing the interactions of different pairs of molecules in a high throughput manner, in choice of different concentration gradients of the GcvB, and with reduced

probing and washing steps. Incorporating these positives assay aspects, a modified version of the co-spotting strategy described in section 4.5.2 was proposed for the next set of experiments to explore disruption of the validated GcvB-mRNA interactions, through addition of an antisense PNA molecule designed to bind to the seed region of GcvB. The PNA was incorporated in the IVT step of the RNA Array protocol and details of the PNA design and disruption assay strategy are detailed below.

5.2.1. Designing a PNA molecule to disrupt GcvB pairing to mRNA targets

As previously observed (Chapter 4, section 4.5.2.3), when the predicted binding region of the target mRNA *ilvC* is scrambled (creating *ilvC*-S), a decrease in the interaction efficiency of GcvB-*ilvC* vs GcvB-*ilvC*-S of approximately 90% is observed. The 5-nucleotide sequence 5' – CACAA – 3' is located in the core of the predicted mRNA binding site for GcvB and repeated three times along the *ilvC* mRNA. This was therefore used as the basis for the PNA sequence design. This A/C-rich sequence is perfectly complimentary to the core portion of the 13-nucleotide sequence of the predicted G/U-rich seed region of GcvB. This suggests optimal GcvB-mRNA target sequence complementarity around this area supporting interaction pairing, and therefore serves as the candidate region for PNA targeting to promote GcvB-mRNA interaction disruption. Based on these observations, and the *in silico* studies from Chapter 3, the inhibitor PNA molecule sequence was designed antisense to the putative G/U-rich seed region of GcvB (Figure 5.3), using a PNA design tool from PNA Bio (Chapter 2, section 2.3.3).

The PNA molecule was synthesised by Panagene Inc. The 10-nucleotide construct included a cell penetrating peptide (CPP) at the 5' end of the PNA. CPPs facilitate cellular intake of a range of molecules, including DNA and NAMs such as PNA. Inclusion of a CPP-tag supports potential future *in vivo* experiments, so will not be considered further as part of this thesis.

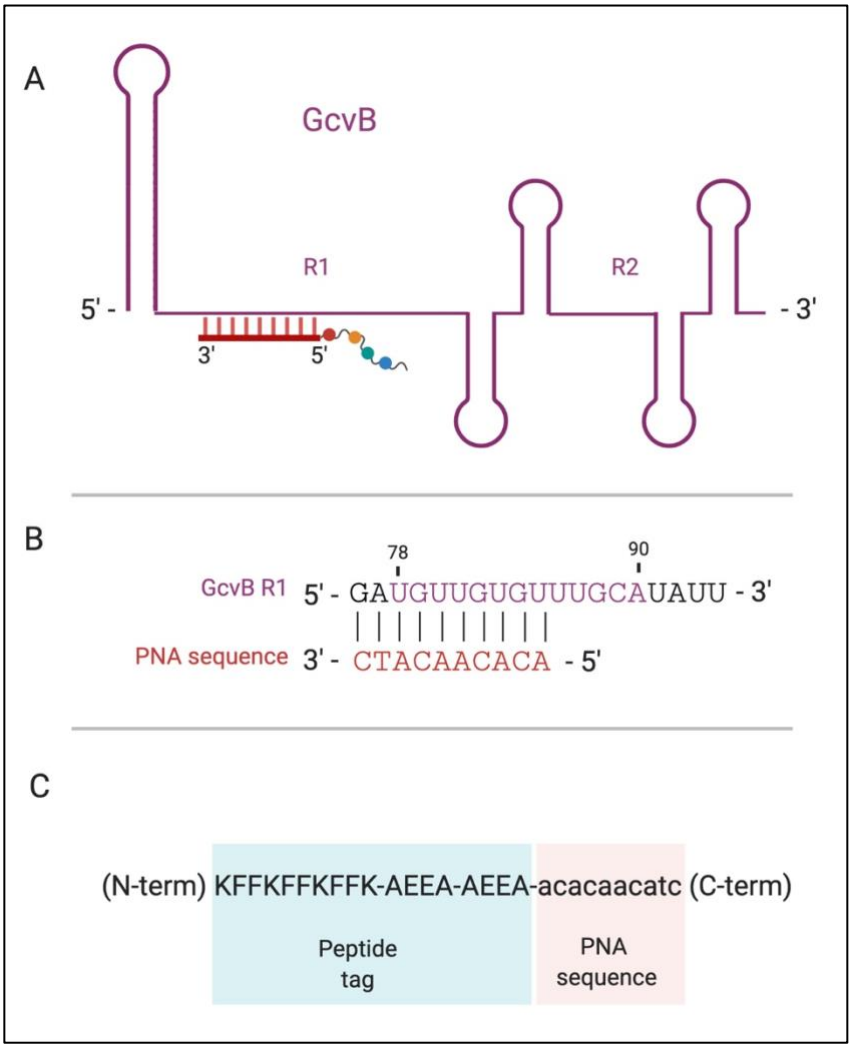


Figure 5.3. PNA sequence design antisense to the G/U-rich seed region of GcvB. A) A schematic of GcvB containing 5 stem-loops and two single stranded conserved regions (R1 and R2). The putative G/U-rich seed region of GcvB is located in R1, and the nine predicted targets mRNAs are anticipated to interact with partial complementarity in this area. The PNA molecule was designed to target this region by antisense complementarity. B) The actual seed region (R1) sequence perfectly paired against the designed PNA sequence. C) The final PNA design with the addition of a cell penetrating peptide. The PNA nucleotide sequence was verified for non-specific binding to mRNA target sequences using the IntaRNA webtool and no interaction was predicted in any functional region of the RNA constructs for all targets.

5.2.2. An experimental strategy for testing PNA disruption of GcvB-mRNA interactions

The co-spotting strategy to exploit the disruption of GcvB-mRNA interactions consisted of mixing the IVT DNA templates of GcvB-MG (often referred to as simply GcvB going forwards in this chapter) and target mRNAs prior to the immobilisation step on the streptavidin coated (SA-coated) slide, as described in the section 4.5.2 of Chapter 4. The standard RNA array protocol was modified to accommodate the addition of the PNA in the

IVT mixture prior to the IVT step, in which SA-coated slides are sandwiched and incubated for 30 minutes at 37^{o}C. However, the addition of PNA to the IVT mixture was expected to disrupt the previously validated GcvB-mRNA interactions, as PNAs present high affinity properties for hybridisation in comparison to RNAs and single stranded DNAs (Summerton, 2006), suggesting the PNA would preferentially bind GcvB instead of an mRNA target. The slides were then separated, and any unbound molecules were removed in a modified washing step, using hybridisation buffer (Appendix 14) instead of the standard protocol of 1 x Phosphate-Buffered Saline (PBS) with 0.02% (v/v) Tween 20 (1 x PBST) (Appendix 14). The washing steps removed molecules unbound to the immobilised mRNAs, specifically GcvB bound to PNA.

Since the RNA slide contains the immobilised mRNAs and any bound GcvB, both of which are unlabelled, probing steps were required to detect both sets of molecules. This involved the addition of two fluorescently labelled antisense ssDNA probes: namely the SA-Linker probe against the SA-Linker of the target mRNA to detect its presence within the spot (described in Figure 4.8, section 4.3.2), and the GAG-Linker probe to detect the presence of GcvB interacting with the immobilised target mRNA (described in Figure 4.27 of section 4.5.2.2). Finally, the probed RNA slide containing the immobilised target mRNAs and any interacting GcvB was analysed to provide information on whether GcvB-mRNA interaction efficiencies had decreased following the addition of PNA. A schematic summary of the strategy is shown in Figure 5.4. Assay configuration and results are presented and analysed in the following sections.

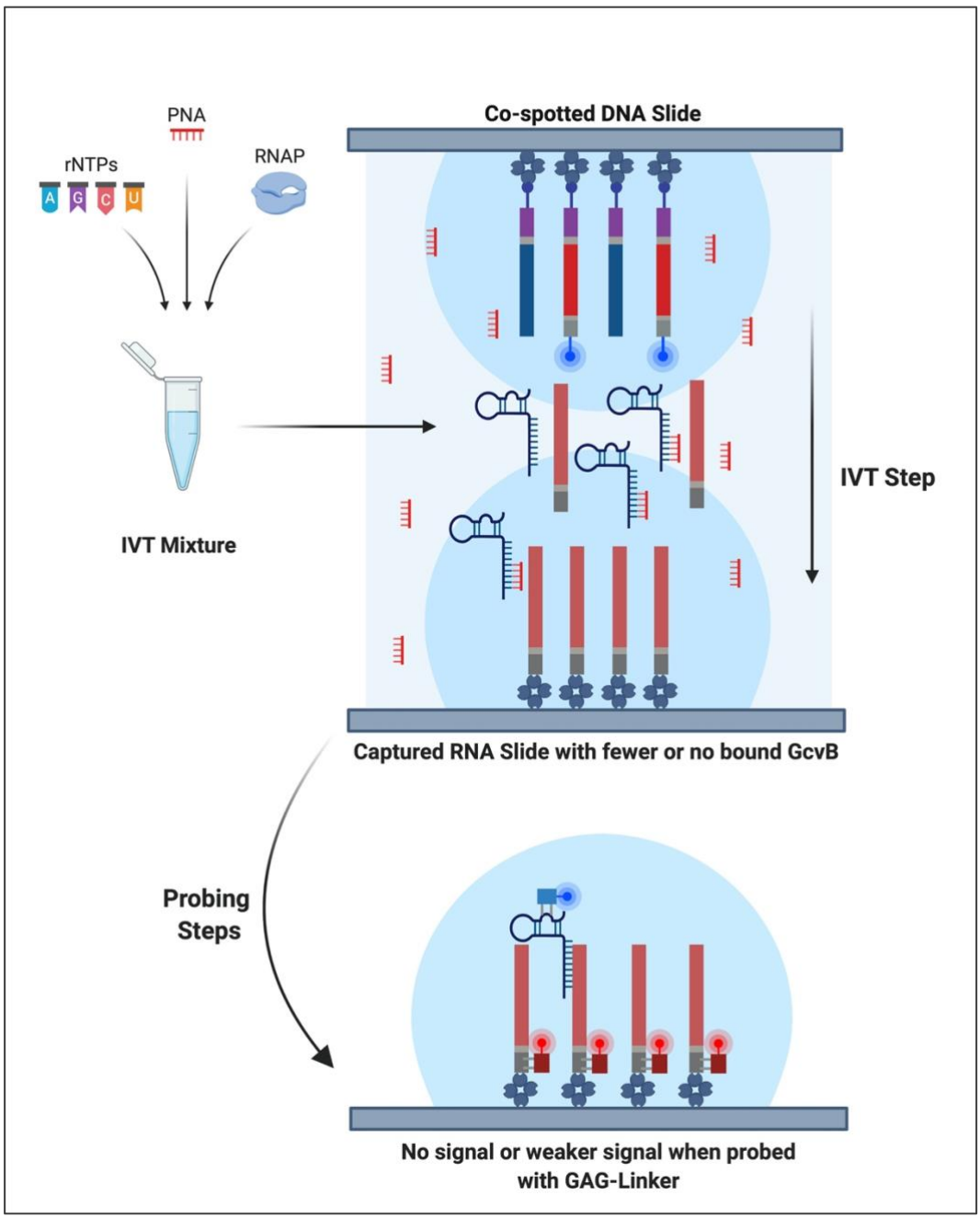


Figure 5.4. Experimental strategy schematic for PNA disruption of a GcvB-mRNA interaction. An SA-coated slide containing the fluorescently labelled (Alexa488 or Alexa647) DNA templates of the target mRNAs and unlabelled IVT DNA templates of GcvB-MG, was used to produce the corresponding unlabelled RNA slide following IVT. The target mRNAs containing the RNA SA Aptamer are captured and immobilised on the SA-coated slide. PNA is free in the IVT mixture and may hybridise with GcvB-MG at a given rate. Transcribed GcvB-MG molecules may bind to PNAs, interact with a target mRNA or remain unbound in the liquid. Unbound molecules are removed during the washing step. The RNA slide is then probed with the GcvB-MG GAG-Linker Alexa488 probe to explore whether any GcvB-mRNA interaction occurred. The signal from the GAG-Linker probe is expected to be weaker following with the addition of PNA, which has been designed to block GcvB pairing to target mRNAs. The SA-Linker Alexa647 probe is also added to the RNA slide to confirm the presence of the mRNA targets in their specific spot, making it possible to quantify interaction efficiencies and to validate the technique.

5.3. Exploring disruption of validated GcvB-mRNA interactions using PNA

Chapter 4 demonstrated successfully using the RNA array approach to investigate GcvB-mRNA interactions. Building on learnings, and as described in 5.2, a PNA has been designed to inhibit GcvB-mRNA pairing through binding to the G/C-rich putative GcvB seed region, and a strategy developed to test PNA impact on the GcvB-mRNA interactions using the RNA array. Details of establishing the PNA-disruption assay on the RNA array, together with the findings and analysis of the impact of PNA on the GcvB-mRNA interactions, are provided below.

5.3.1. Establishing a PNA-disruption assay on the RNA array

Using the experimental strategy in section 5.2.2, DNA templates of GcvB-MG and the mRNAs were synthesised with the 5' biotinylated immobilisation oligo, but only the target mRNA DNA templates were synthesised with 3' fluorescent Alexa488 label. Because template spotting concentrations over 200 nM may saturate the spot for some DNA IVT templates, the final concentration of each molecule was set to 100 nM to accommodate for co-spotting. Two DNA slides were generated by co-spotting the DNA templates with an identical slide configuration, as illustrated schematically in Figure 5.5.

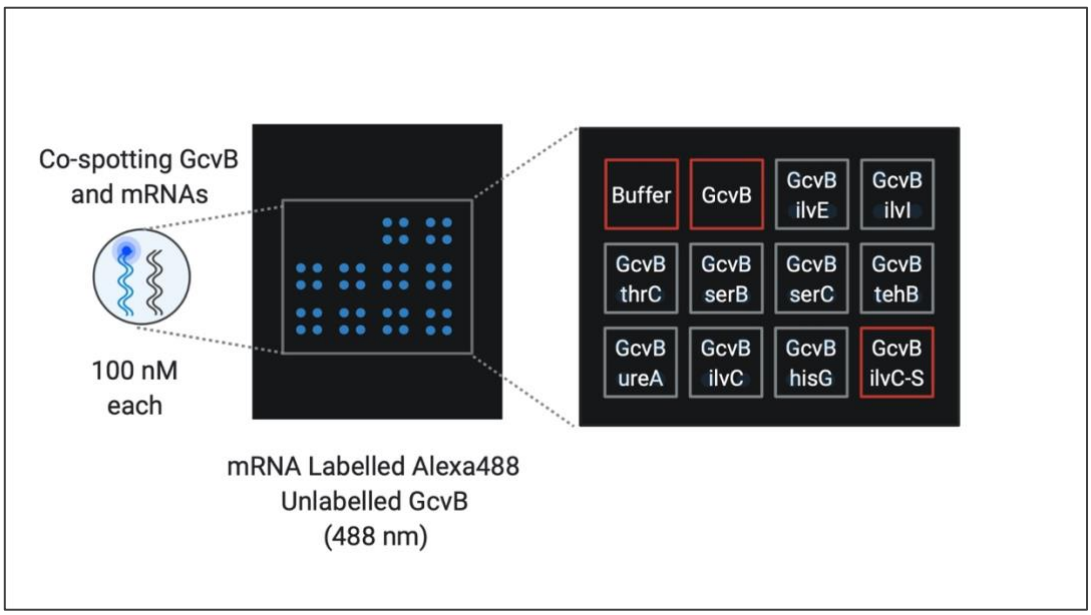


Figure 5.5. DNA slide spotting schematic for the PNA-disruption assay. The SA-coated DNA slide contained automated high throughput spotting of IVT DNA templates: each field accommodated a four-spot repetition of co-spotted IVT DNA templates for GcvB-MG and the nine predicted target mRNAs plus *ilvC-S,* with final constant concentration of 100 nM each. The IVT DNA templates for the targets were fluorescently labelled with Alexa488, while GcvB-MG templates were unlabelled. Two control fields were added, one with buffer only and the other was reserved for the GcvB-MG DNA template on its own.

Once the DNA templates were immobilised, one of the DNA slides was sandwiched with its partner RNA slide with a standard IVT mixture recipe, without the PNA molecule. The remaining DNA slide was also sandwiched with its partner RNA slide, but the IVT mixture was complemented with 500 nM of the previously designed PNA molecule. The choice of concentration for the inhibitory molecule was based on the standard concentration used for the added detection probes. Both sandwiched slide pairs were incubated in a humidified chamber following the standard RNA array protocol, and then transferred to room temperature for extra 30 minutes to allow for GcvB interaction with its target mRNAs. The slides were then separated and washed with hybridisation buffer. The slides with the captured RNA and any interacting GcvB were incubated with 500 nM of the GAG-Linker Alexa 488 probe, against the extra linker in GcvB (i.e., GcvB-MG), and with the SA-Linker probe Alexa647 antisense to the SA linker in the mRNA constructs (Chapter 4, section 4.5.2). Both slides were then visualised using excitation wavelengths of 488 nm and 635 nm with Standard Blue and Standard Red emission filters, respectively (Figure 5.6).

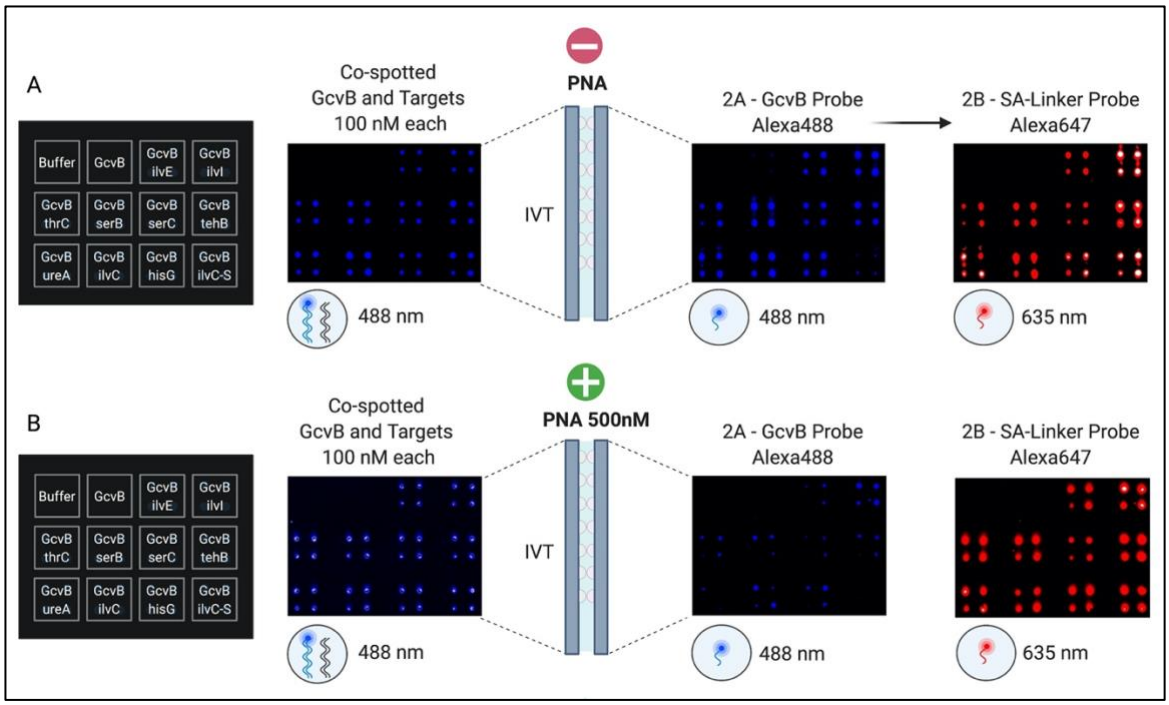


Figure 5.6. Visualisation of the PNA-disruption assay results. Actual visualisation of slides by the GenePix scanner. A) Schematic of the DNA IVT template slide layout is shown adjacent to visualisation of the fluorescently labelled templates on the slide. The RNA array is generated without the addition of PNA, probed with the GcvB-MG GAG-Linker probe Alexa488 visualised at 488 nm (2A) and the SA-Linker Alexa647 probe at 635 nm (2B). B) Schematic of the DNA IVT template slide layout and visualised slide is as in (A). The RNA array is produced with the addition of PNA in the IVT mixture, and the slide visualised following identical probing steps to those described in A.

Visual analysis of the RNA slide generated without the addition of PNA, scanned at 635 nm to detect the mRNA-bound SA-Linker probe, exhibited signal for all of the nine target mRNAs and *ilvC*-S. This confirmed successful immobilisation of the mRNAs on the RNA slide. The same RNA slide, scanned at 488 nm, confirmed the presence of GcvB, through identification of the bound GAG-Linker probe indicating GcvB-mRNA interaction for all of the nine target ROIs, excluding the negative control *ilvC*-S, as expected. These results are consistent with the findings in Chapter 4, and for this reason, further detailed analysis of interaction studies is not provided in this section. Importantly, the control fields spotted with buffer, and the IVT DNA template for GcvB-MG, displayed no notable fluorescence signal for both wavelengths, indicating that no non-specific binding of GcvB-MG or spotting contamination had occurred. By contrast, for the RNA slide generated with the addition of PNA in the IVT mixture, scanned at 635 nm to detect the mRNA-bound SA-Linker probe exhibited signals for all the nine target mRNAs and *ilvC*-S, whilst scanning at

488nm revealed that fluorescent signal for the GcvB-MG-bound GAG-Linker probe was substantially reduced. This suggested that the addition of PNA may have reduced the interactions between GcvB and its target mRNAs, possibly as a consequence of the PNA binding to GcvB and thereby disrupting its ability to pair with its partner mRNA. This evidence of PNA-disruption of GcvB-mRNA interactions, generated from a modified RNA array assay, will be expanded and explored further below.

5.3.2. Disrupting GcvB-mRNA interactions with PNA

Section 5.3.1 established the PNA-disruption assay on the RNA array and demonstrated that the addition of 500 nM PNA to the IVT mixture in the IVT step of the RNA Array protocol affects GcvB-mRNA pairing for all the nine mRNA targets. However, under the conditions tested, it was seen that concentrations higher than 500nM would be required to completely inhibit GcvB-mRNA pairing. To explore this, it was necessary to undertake further experiments in which higher PNA concentrations were tested. The DNA IVT template slide layout is shown in Figure 5.7 and experiments were carried out as in 5.3.1, except for that 4 slides were prepared to allow testing of 0 nM, 500 nm, 750 nM and 1000 nM PNA concentrations. The maximum concentration of 1000 nM PNA was chosen as it represents double the final concentration of both probes (GAG-Linker for GcvB-MG and SA-Linker for the target mRNAs) and was therefore thought suitable for array studies. Additionally, a number of repeated technical controls were removed or reduced, as it has been consistently shown that GcvB-MG specifically binds to the immobilised mRNAs, without diffusing to other parts of the slide.

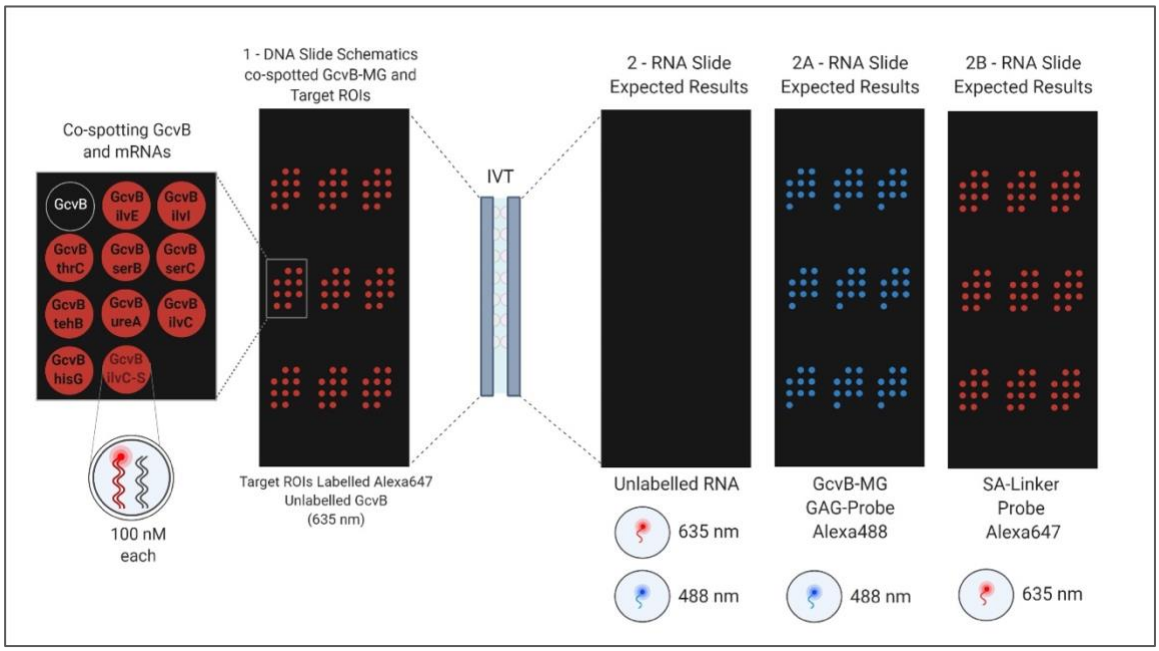


Figure 5.7. DNA slide spotting schematic for the PNA-disruption assay and associated RNA array details. The SA-coated DNA IVT template slide (1) contains nine fields. Each field accommodates 11 spots in total representing the IVT DNA templates for each of the nine target mRNAs and *ilvC*-S co-spotted with GcvB-MG, plus a spot for the DNA template of GcvB-MG spotted on its own as a control monitoring for non-specific binding of GcvB-MG. The templates were spotted with a final constant concentration of 100 nM each. The IVT DNA templates for the target mRNAs and *ilvC*-S were fluorescently labelled with Alexa647, while GcvB-MG templates were unlabelled. Following IVT, the unlabelled RNA slide (2) is then probed with (2A) GAG-Linker probe labelled with Alexa488 which is visualised at an excitation wavelength of 488 nm using a standard blue emission filter and indicates GcvB presence on the RNA array, and (2B) the SA-Linker probe labelled with Alexa647 which is visualised at an excitation wavelength of 635 nm using a standard red emission filter and identifies immobilised mRNAs.

Visual inspection of the data produced (Figure 5.8), indicated a significant reduction in fluorescence signal for the GAG-linker probe at 488 nm, which identifies bound GcvB, aligned with an increasing concentration of PNA within the IVT step of the RNA array protocol. Specifically, in the absence of PNA (Figure 5.8, A), all co-spotted target mRNAs, with exception of the negative control *ilvC*-S, showed signal indicating interaction with GcvB to immobilised target mRNAs, identified by the SA-Linker probe when visualised at 635 nm. For the same experiment in the presence of 500 nM, 750 nM, and 1000 nM PNA, immobilised target mRNAs were identified by the SA-Linker probe when visualised at 635 nm. However, the level of bound GcvB was reduced for the 500 nM (Figure 5.8, B) and 750 nM (Figure 5.8, C) PNA conditions, identified by the reduced signal following probing with the GAG-linker probe and scanning at 488 nm. This suggests the PNA has partially disrupted GcvB from pairing to its partner mRNAs. When PNA was present at 1000 nM (Figure 5.8,

D), no GcvB signal was identified, suggesting the complete disruption of GcvB pairing to its mRNAs. In line with previous results, no significant evidence of GcvB interacting with immobilised *ilvC*-S was obtained for any of the conditions tested.

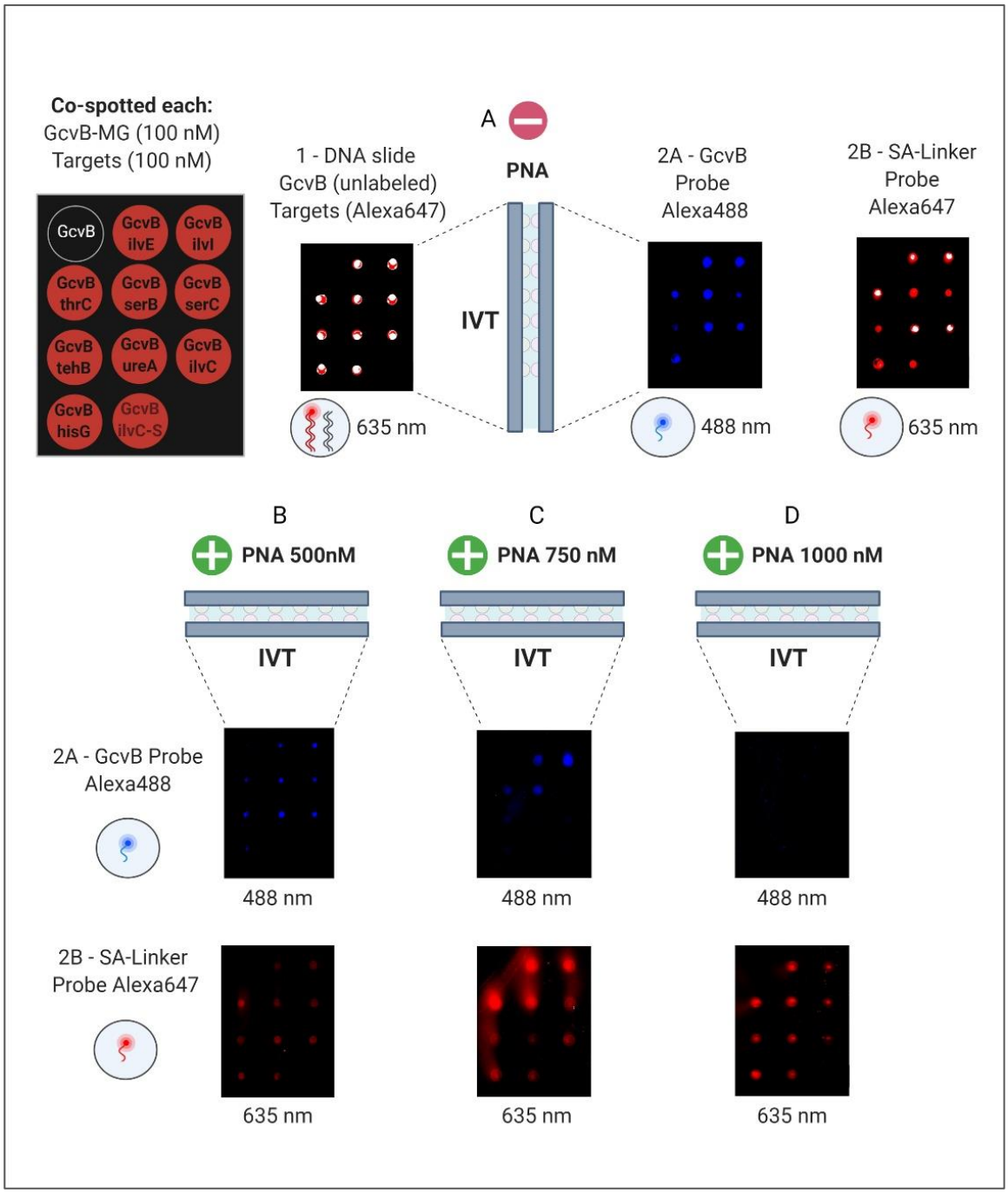


Figure 5.8. Visualisation of the results from the PNA-disruption assay. Actual visualisation of slides by the GenePix scanner: four identically spotted DNA slides (1) were used to generate the corresponding RNA array on the RNA capture slides. A) RNA array generated without the addition of PNA in the IVT step of the RNA Array protocol, probed with the GcvB-MG GAG-Linker probe Alexa488 visualised at 488 nm (2A) and the SA-Linker Alexa647 probe at 635 nm (2B). This strategy was repeated for the other three slides generated with the addition of 500 nM (B), 750 nM PNA (C) and 1000 nM PNA in the IVT mixture (D).

The mean fluorescence for the bound GcvB was normalised for level of immobilised mRNA, for each PNA condition tested, and the interaction efficiencies displayed graphically in Figure 5.9, A. The data shows that in the absence of PNA, in line with previous findings, GcvB specifically interacts with all nine targets with differing levels of efficiency, except for the control, *ilvC*-S, for which no significant binding was seen and so has been omitted from the graph. As the level of PNA increases to 500 nM and 750 nM, the interaction efficiency is reduced, such that by 1000 nM PNA, there is no notable GcvB-mRNA interaction evident.

Considering the disruption efficiency of the PNA for each of the GcvB-mRNA interactions, which identifies the change in interaction efficiency upon PNA addition, GcvB-*ilvI* is most impacted. With the highest interaction efficiency, the presence of 500 nM PNA disrupts this by 63%, compared to GcvB-*serC* which has one of the lowest interaction efficiencies and is only disrupted by 30% by the presence of PNA, under the same conditions. Disruption efficiencies for 500 nM PNA show the most difference between the various GcvB-mRNA interactions (Figure 5.9, B). This can be considered as reflecting differences between the interaction efficiencies of the pairings themselves (Chapter 4, section 4.4.3.2), the binding preferences of GcvB to the PNA vs the mRNA, the possible differences in the region of GcvB that pairs to each mRNA and the associated impact of whether or not the PNA binding the GcvB completely or partially blocks the pairing site. These differences are not apparent for the 750 nM and 1000 nM PNA concentrations, which result in disruption efficiencies around the 80% and 90% levels respectively for all GcvB-mRNA pairings.

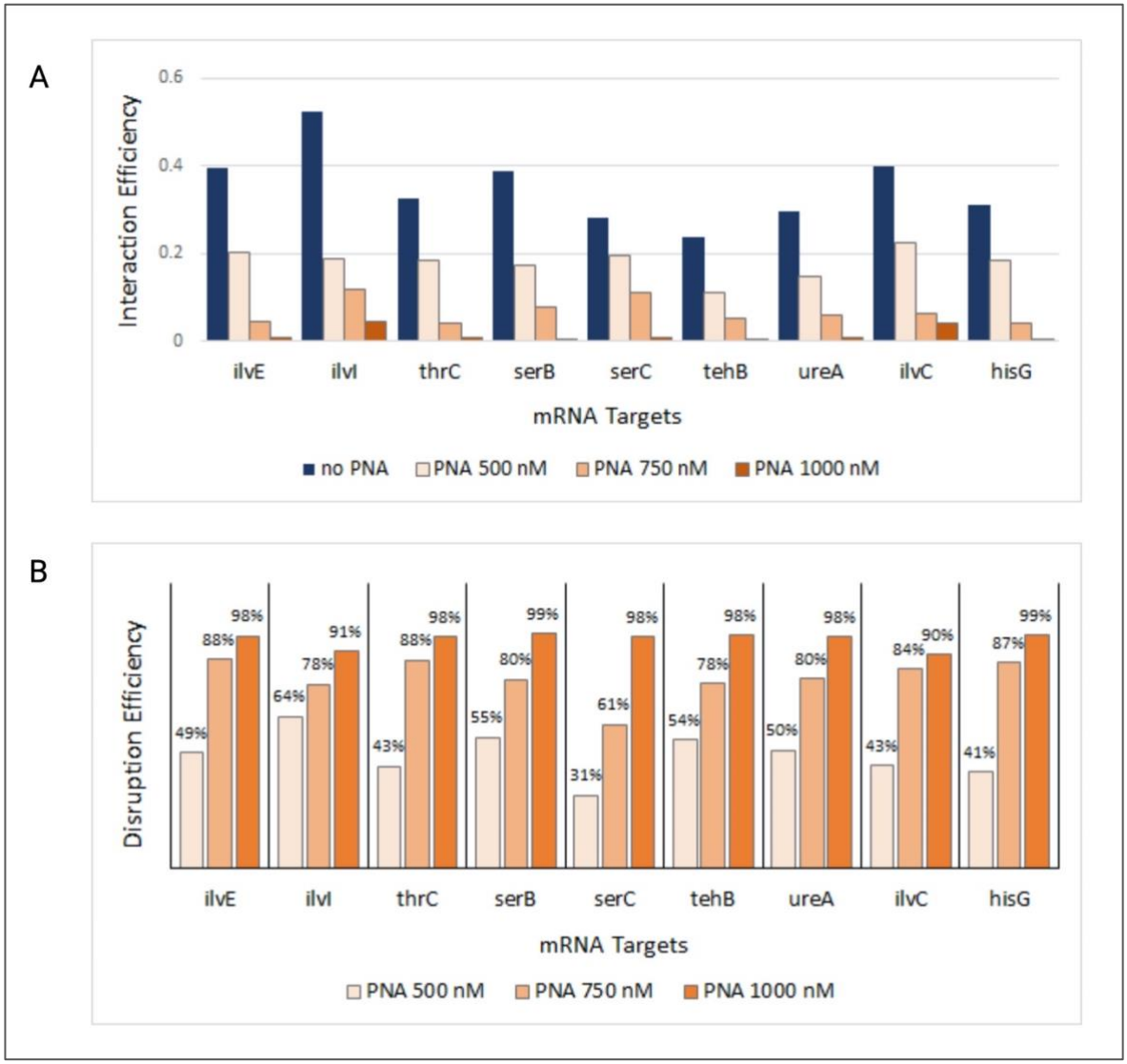


Figure 5.9. Analysis of PNA disruption of GcvB-mRNA pairing. The figure contains two graphs; A) the graph shows the interaction efficiency of GcvB bound to the immobilised target mRNAs in the presence of PNA concentrations from 0-1000 nM, represented by dark blue bars (0 nM PNA) and orange bars coloured in three tones, light (500 nM PNA), moderate (750 nM PNA) and dark (1000 nM PNA) shades. B) The graph plots the disruption efficiency for GcvB and each target mRNA, for each concentration of PNA, and represents the interaction efficiency disrupted as a percentage of the total seen for the GcvB-mRNA interaction in the absence of PNA. The colour scheme is the same as that used for the upper graph.

5.3.3. Exploring PNA-disruption of specific GcvB-mRNAs interactions in more detail

Section 5.3.2 identified that the PNA designed to bind to the putative G/U-rich seed region of GcvB was capable of disrupting all GcvB-mRNA interactions tested. By comparison, section 4.5.2.3. allowed a more detailed assessment of GcvB pairing to specific mRNAs, namely *ilvI*, *ilvC* and *hisG*. These two approaches were combined to explore, in

more detail, disruption of GcvB interactions with *ilvI, ilvC* and *hisG* as a consequence of PNA presence. The DNA IVT template slide schematic is shown in Figure 5.10 and consisted of a concentration gradient of GcvB-MG co-spotted with a constant concentration of either ilvI, *ilvC*, *hisG* or *ilvC*-S as a control. The experiment was carried out as in 5.3.1, except that three slides were generated to allow testing of 0 nM, 500 nm, and 1000 nM PNA concentrations. As previously, the RNA slide was probed with the GcvB GAG-Linker probe to determine the presence of GcvB, and with the SA-Linker Alexa647, to quantify the amount of immobilised target mRNAs.

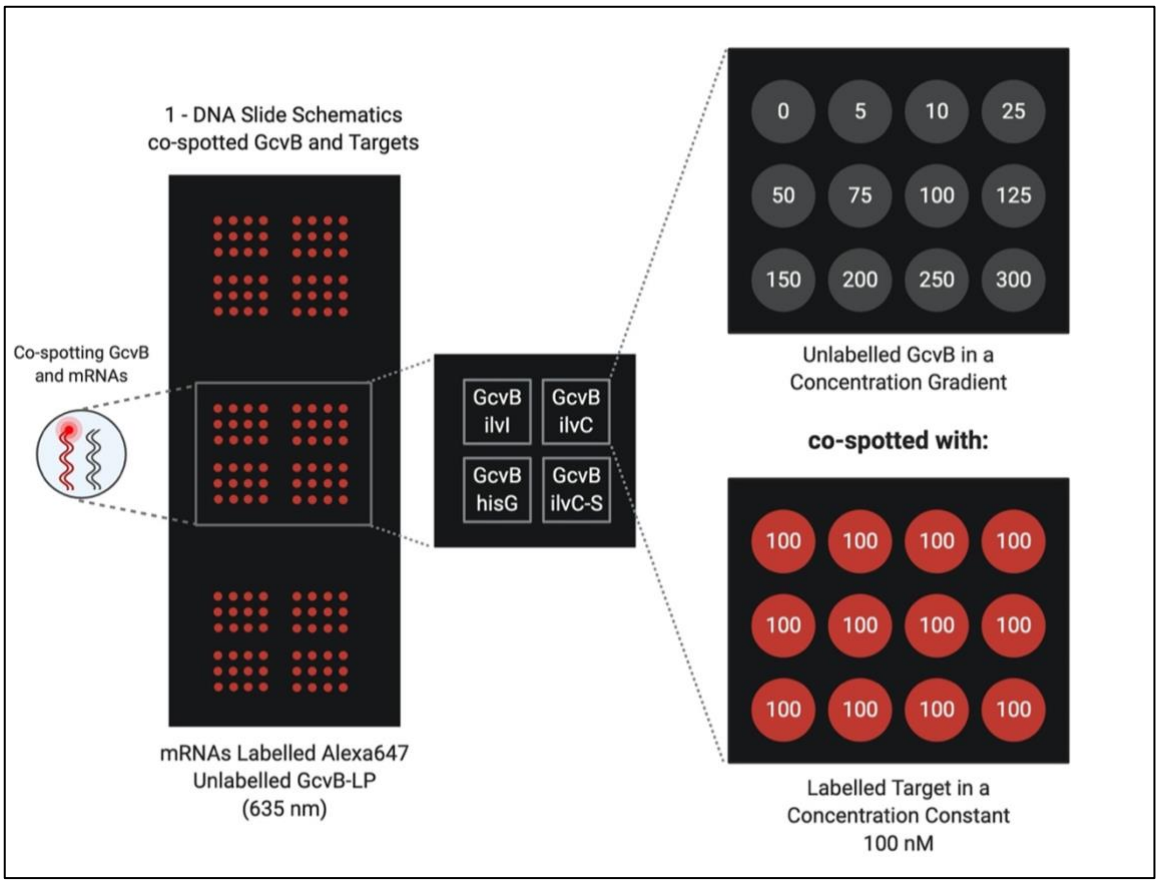


Figure 5.10. DNA slide spotting schematic for the PNA-disruption assay involving co-spotting different concentrations of GcvB-MG with a set concentration of target mRNA. The spotting configuration of the mixed DNA IVT templates was set in a 4 x 3 pattern for a selection of three targets (*ilvI, ilvC* and *hisG*) and *ilvC*-S (control) only. The pattern was repeated three times in the same slide to generate replicates. The IVT DNA template for GcvB-MG was added to the co-spotting sample in a 16-spot gradient concentration ranging from 0 nM to 300 nM, while the concentration for the DNA IVT template of the target mRNAs remained constant at 100 nM.

Visual inspection of the resulting data once again indicated that PNA disrupts GcvB pairing to mRNA targets *ilvI*, *ilvC* and *hisG* (Figure 5.11), as seen by the loss of probe signal indicative of GcvB presence, compared to mRNA probe levels, for experiments with PNA present compare to when PNA was absent.

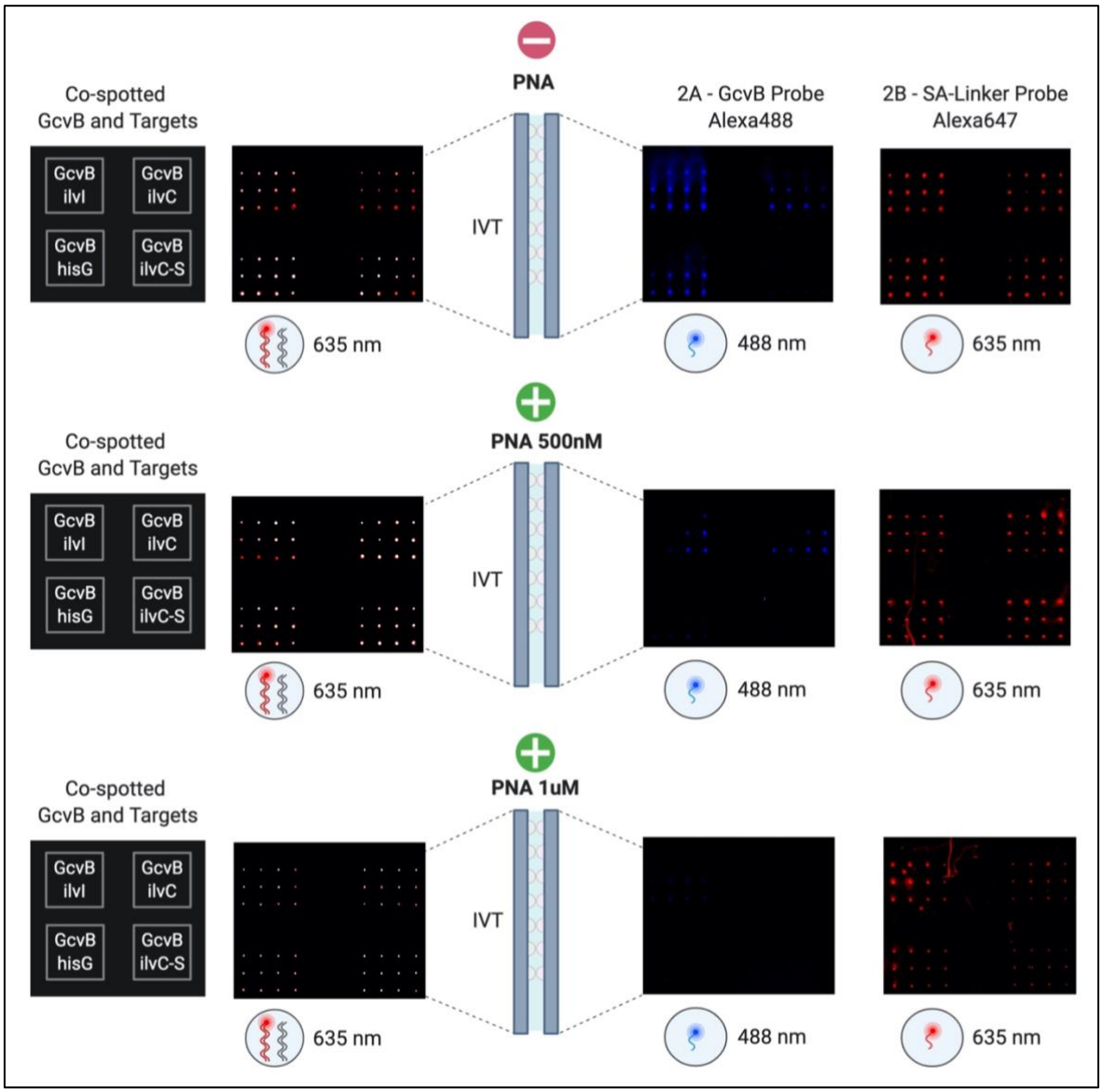


Figure 5.11. Visualisation of the results exploring disruption of specific GcvB-mRNA interactions with PNA in more detail. Actual visualisation of slides by the GenePix scanner: three identical DNA IVT template slides co-spotted with IVT DNA templates of the target mRNAs labelled with Alexa647 and unlabelled GcvB-MG were used to generate the corresponding RNA array on the RNA capture slides. A) RNA array generated without the addition of PNA, probed with the GcvB-MG GAG-Linker probe Alexa488 visualised at 488 nm (2A) and the SA-Linker Alexa647 probe at 635 nm (2B). This strategy was repeated for the other two slides generated with the addition of 500 nM PNA (B) and 1000 nM PNA (C).

The fluorescence of the bound GcvB was normalised for the level of immobilised mRNA for each different co-spotted GcvB concentration and PNA condition tested, and the interaction efficiencies displayed graphically in Figure 5.12. Logarithmic trendlines have been added as a visual aid only. The data shows that in the absence of PNA, for increasing concentrations of co-spotted GcvB, the interaction efficiency of GcvB to the immobilised mRNA target increased with increasing concentration of GcvB-MG DNA IVT template, before reaching a plateau. This trend applied to all three mRNAs tested, with *ilvI* having higher interaction efficiencies for all GcvB concentrations tested, compared to *ilvC*, and *hisG* having the lowest interaction efficiency for all GcvB concentrations tested. A similar profile was observed when 500nM PNA was included in the assay, although for each mRNA, the interaction efficiency was lower for all GcvB concentrations tested compared to in the absence of PNA. Finally, in the presence of 1000nM PNA, only minimal interaction efficiencies were seen for all GcvB concentrations tested.

Considering each mRNA target, addition of 500nM PNA affects the interaction efficiencies differently. For example, 500 nM of PNA approximately halves the interaction efficiency of GcvB-*ilvI* and GcvB-*hisG* whereas it only reduces it by a third for GcvB-*ilvC*. In comparison addition of 1000nM PNA similarly affects the interaction efficiencies for all GcvB-mRNAs, reducing them by approximately 90% or more. This starts to give insights into the comparative affinities of the interactions, indicating differences in the competition of GcvB for PNA vs *ilvI, hisG* or *ilvC*. Whilst PNA similarly affects the GcvB-*ilvI* and GcvB-*hisG* interactions, GcvB seems to have more preference for *ilvC* binding compared to PNA, as the PNA does not reduce GcvB-*ilvC* interaction efficiency by as much. This provides a basis for informing future experiments to explore PNA-disruption of GcvB-mRNA interactions in more detail going forward.

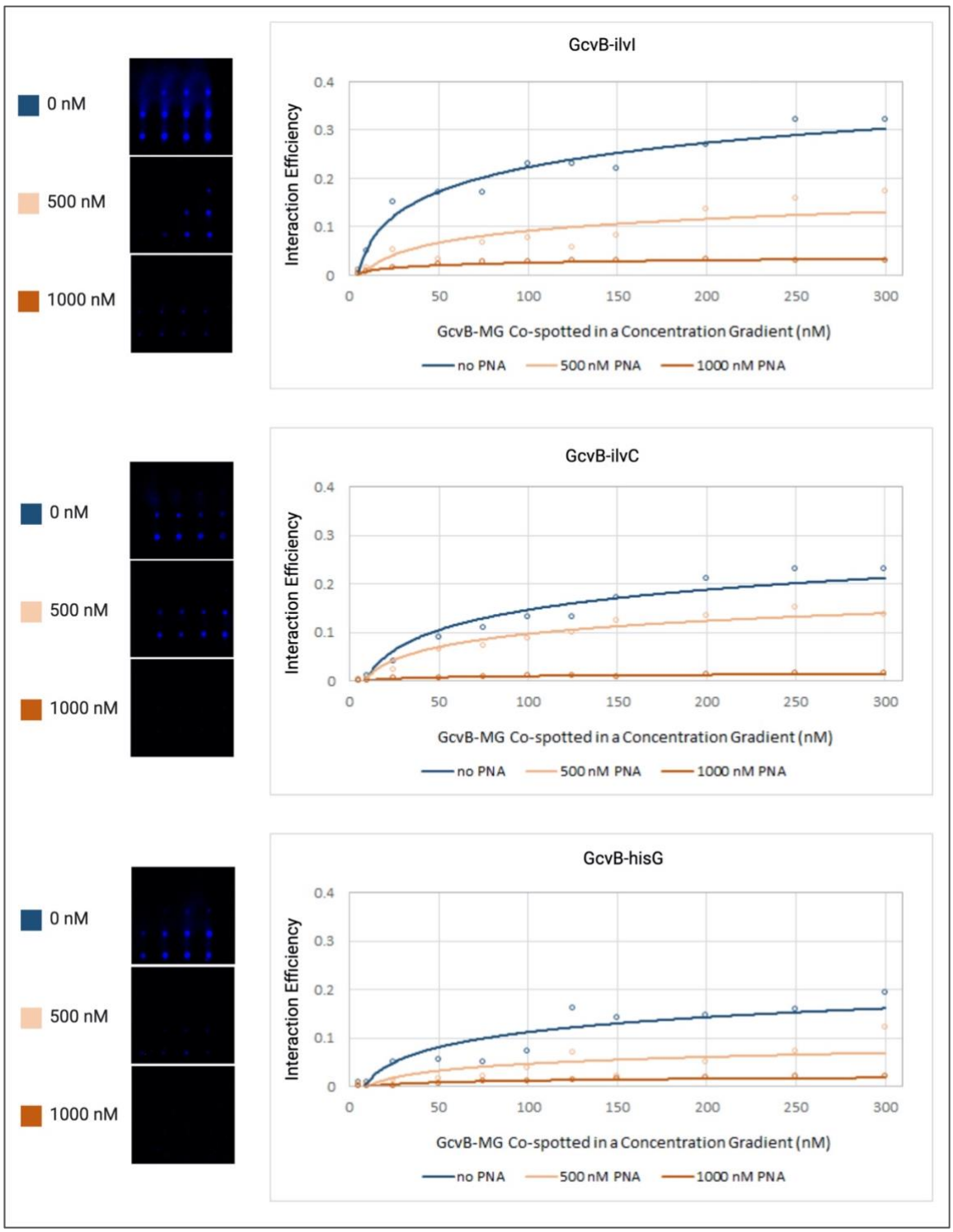


Figure 5.12. Analysis of the results exploring disruption of specific GcvB-mRNA interactions with PNA in more detail. The graphs show the analysis of the interaction efficiency of GcvB-MG bound to the immobilised target mRNAs when PNA is added at different concentrations, represented by orange bars coloured in two tones, light (500 nM) and dark (1000 nM) shades. The blue bars represent the 0 nM PNA control. The interaction efficiencies are plotted for the increasing concentrations of co-spotted GcvB-MG. Logarithmic trendlines have been added as a visual aid only

5.4. Summary

Building on the work in Chapter 3, coupled with insights on the GcvB-mRNA interactions from the studies in Chapter 4, this chapter has focused on the design and testing of a PNA to specifically disrupt GcvB-mRNA interactions. The PNA was designed antisense to the putative G/U-rich seed region of GcvB. A PNA-disruption assay was established on the RNA array. This built on the GcvB/mRNA co-spotting approach developed in Chapter 4 but incorporated the PNA within the IVT step of the RNA array protocol. Data indicated that the PNA successfully disrupted GcvB-mRNA interactions efficiencies to varying levels at the lower concentration tested, and almost completely disrupted the interactions at the higher concentrations tested.

Work in Chapter 3 had noted that the predicted hybridisation energies for the pairing of GcvB to its mRNA targets vary, recognising that GcvB-*ilvC* and GcvB-*ilvI* have lower hybridisation energies than GcvB-*hisG*, for example (Table 3.3; Appendix 7). Such differences can partially be explained by the degree of pairing between GcvB and its mRNA targets (Figure 3.12). These predictive results aligned with the experimental findings in Chapter 4 where different interaction efficiencies were seen for the GcvB-mRNA target pairs, and with *ilvI* and *ilvC* repeatedly being identified as having the highest interaction efficiencies compared to GcvB-*hisG* as well as the other mRNA targets (Figure 4.32 and Figure 4.35).

Given these differences, it is unsurprising that the addition of PNA results in different disruption efficiencies for the various GcvB-mRNA pairs. For example, GcvB-*ilvC* hybridisation involves a region of 18 consecutive nucleotides, whereas for *ilvI* it involves 10 nucleotides. With a PNA of 10 nucleotides binding to a region that only partially disrupts these pairing sites, it is therefore not unexpected that the PNA has less disruptive effect on GcvB-*ilvC* (it reduces it by ~30% when added at 500nM), which has a larger region involved in hybridisation pairing, when compared to GcvB-*ilvI* which is more significantly disrupted (by ~50% for 500nM PNA) and has a smaller hybridisation site (Figure 5.13).

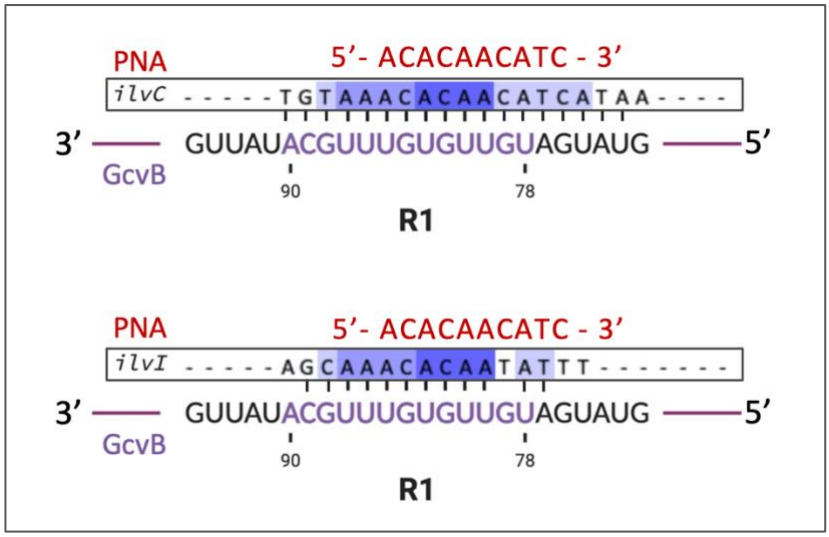


Figure 5.13. Schematic illustration of the PNA aligned with the predicted pairing site for GcvB hybridised with *ilvC* and *ilvI*. The illustration shows the GcvB sequence within the R1 region seed region hybridised to either *ilvC* (upper) or *ilvI* (lower). The percentage identity for the mRNAs was coloured in purple gradient. The PNA sequence (red) is aligned to where it is expected to pair with GcvB, in place of the partner mRNA.

Whilst focusing on pairing sites and hybridisation energies can be informative, other aspects, such as RNA structure, can also come in to play in terms of determining the strength of an interaction between GcvB and its mRNA target. However, considering the hybridisation energies and size of the GcvB-mRNA pairing sites, together with details of the PNA sequence and it's predicted binding site on GcvB, can help towards understanding trends observed. This is important for informing improvement strategies for PNA design with the aim of enhancing PNA disruption for reduced PNA concentrations. With the PNA-disruption assay established for the RNA array, testing the GcvB-mRNA pairs with alternative PNA designs is now possible.

Chapter 6

6. Summary and Perspectives

6.1. Background

Although transcriptional gene regulation controls RNA transcript production, the ability of tightly regulate existing transcripts is highly important for unicellular organisms, such as bacteria, in order to conserve energy to survive and thrive in rapidly changing environments. Post-transcriptional gene regulation (PTGR) supports energy trade-off throughout all protein synthesis processes (transcription, translation and decay of mRNAs and proteins) by promoting a prompt shift of energy expenditure and nutrients to more urgent or essential processes in the cell in response to environmental changes (Hausser *et al.*, 2019). However, protein translation is the most expensive anabolic process for a growing unicellular organism, as it is estimated to use up to 70% of the ATP budget (Stouthamer, 1973; Pontes *et al.*, 2015). For example, to produce an average single protein containing 300 amino-acids, translational machinery may use approximately 1350 ATP and GTP molecules (for amino-acid incorporation and enzymatic activity), 1650 carbon and 540 nitrogen atoms (Stouthamer, 1973; Kafri *et al.*, 2016; Hu *et al.*, 2020). Thus, halting translation of mRNA transcripts into protein seems an advantageous strategy to conserve energy as quickly as possible when the protein is no longer required.

There is growing evidence that bacterial non-coding small RNAs (sRNAs) play a major role in PTGR, controlling key metabolic pathways which, in many cases, are directly linked to bacterial virulence (Waters & Storz, 2009; Vogel & Luisi, 2011; Holmqvist *et al.*, 2012). Thus, the goal of this research was to investigate novel *trans*-acting sRNAs linked to pathogenicity in the Gram-negative bacterium *Actinobacillus pleuropneumoniae* (APP), the causative agent of porcine pleuropneumoniae, for therapeutic targeting purposes. The research findings and outputs achieved as part of this study, and the steps that future work

could take, will be summarised below, before considering the work within the broader context of the field.

6.2. A bioinformatic pipeline identifies novel *trans*-acting sRNAs in APP relevant to virulence

In Chapter 3, a selection of *trans*-acting sRNA candidates in APP serovar 8 MIDG2331 and their putative mRNA targets were produced. This was accomplished by the creation of a bioinformatics pipeline based on a combination of existing computational approaches utilising genomic and RNA-seq data from MIDG2331 provided by collaborators from Imperial College London. The collaborators had previously developed an initial computational strategy to predict sRNAs in bacteria (Rossi *et al.*, 2016). However, as the current tools predict *cis* and *trans*-acting sRNAs alike, the pipeline was organised to chronologically mine and scrutinise prediction data by using conditional parameters to select for *trans*-acting sRNAs only.

The result was a selection of nine novel sRNA candidates in APP MIDG2331 curated from RNA-seq data. From this, homology studies identified the predicted sRNA ARRC01 as the putative sRNA GcvB in APP serovar 8 MIDG2331 (GcvB), a previously predicted novel sRNA in APP by collaborators (Rossi *et al.*, 2016) and widely studied in *E. coli* and *Salmonella*. Surprisingly, none of the other predicted novel sRNAs in the list were identified in *Pasteurellaceae*.

The prediction of mRNA targets of GcvB in APP was also generated from the RNA-seq data, resulting in a selection of nine putative mRNA partners, of which seven are involved in amino-acid biosynthesis (*ilvE, ilvI, ilvC, thrC, serB, serC*, and *hisG*), one is associated with urea degradation (*ureA*), and the final one is associated with oxidative stress resistance (*tehB*). The final output generated was therefore a list of potential mRNA partners to the sRNA putatively identified as GcvB. Given the established structural profile of GcvB and its predicted mRNA targets being linked to metabolic pathways associated with bacterial pathogenicity, the GcvB-mRNA target pairs were viewed as ideal candidates for taking forward for further study.

Using the RNA array technology, these interactions were therefore experimentally tested (Chapter 4; and considered below in 6.3), following appropriate establishment of a suitable RNA array methodology. Further, with links to pathways involved in bacterial virulence, a PNA specific to the GcvB seed region was designed and its ability to disrupt GcvB-mRNA interactions was tested using an RNA array PNA-disruption assay (Chapter 5; and considered below in 6.4).

Finally, the remaining list of novel sRNA candidates identified from the APP RNA-seq data form the basis for future *in silico* studies using the bioinformatics pipeline developed here to identify mRNA partners. These can then be taken forward for experimental sRNA-mRNA interaction and associated PNA-disruption studies. Importantly, the bioinformatics pipeline itself also provides a valuable tool for identification of *trans*-acting sRNAs, and predicting their mRNA partners, using RNA-seq data for other bacteria of interest, including "ESKAPE" bacteria, which comprise *Enterococcus faecium*, *S. aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *P. aeruginosa*, and *Enterobacter spp* (Rice, 2008). These are antibiotic resistant and major worldwide life-threatening pathogens, requiring therefore, development of novel antibacterial solutions (reviewed in Felden & Cattoir, 2008).

6.3. The RNA Array technology verifies predicted GcvB-mRNA interactions

The RNA array technology is a novel approach developed by the Callaghan group at the University of Portsmouth. It can be widely applied to biophysically characterise RNA interactions in a high throughput (HTP) manner. It is particularly relevant to investigating post-transcriptional gene regulation by sRNAs in bacteria due to them having a wide and varied number of mRNA targets (Chapter 1, section 1.5 and Chapter 2, section 2.2.1). Specifically, the numerous mRNA targets can be simply generated on an RNA array in one step, and then probed all at once with the sRNA target, thus experimentally identifying many mRNAs interacting with a single sRNA in one experiment.

The RNA array standard protocol was adapted to accommodate the specific requirements of this research. In this study, the technique was used to repeatedly

immobilise numerous distinct mRNAs on the same slide, for probing with a single interaction partner molecule (sRNA) with associated detection probing steps to confirm surface-bound and interacting molecules (Chapter 4, section 4.5.1 and 4.5.2).

Although the RNA array standard protocol is robust and has been previously modified by colleagues (Phillips *et al*., 2018; Norouzi *et al*., 2019; Henderson *et al*., 2019) strategies were further adapted in this work. For example, the number of different mRNAs tested made it possible to analyse transcription efficiencies among a wider range of IVT DNA templates than has been possible previously. Although all IVT DNA templates of the target mRNAs contained a G(GG) sequence to ensure that there is a G in positions +1, +2 and +3, as required for optimal transcription efficiency using T7 (Imburgio *et al.*, 2000), the results showed that there was a difference in relative transcription efficiency across the templates. Whilst the total amounts of transcribed RNA may differ among experiments and technical repeats for the same template, there is a constant relative pattern of transcription efficiency by template.

For instance, the IVT DNA template for the target mRNA *serC* has consistently displayed the lowest transcription efficiency in comparison to the other templates in all performed assays, while RNA yield produced by the DNA template of the target mRNAs *ilvE* and *ilvI* has been consistently high in comparison (Chapter 4, section 4.4.3.1). With increased data for a range of additional targets, and computational analysis to mine transcription efficiencies relative to nucleotide sequence, future work could allow it to be possible to explore further if there is a link between T7 transcription efficiency and sequence requirements in addition of what is already known about the preference for G in positions +1, +2 and +3. The RNA array is ideally suited to generating this data, as the transcription efficiencies of many transcripts can be output in one experiment for subsequent assessment.

Regarding the interaction studies, GcvB molecules were synthetically altered to accommodate a strategy for detection of binding. This included either internally Cy-labelling GcvB for direct probing studies (Chapter 4, section 4.5.1), or modifying GcvB (i.e., GcvB-MG) such that it contained a specific linker sequence for subsequent probing (Chapter 4, section 4.5.2), as used in the co-spotting studies. Both strategies were shown to demonstrate GcvB-mRNA interactions for the nine mRNA targets. However, whilst

interaction efficiencies varied between GcvB and the mRNA targets, consistent patterns emerged suggestive, and in line with, the hybridisation energy trends noted for pairing predicted as part of the *in silico* studies in Chapter 1. For example, interaction between GcvB-tehB has regularly produced lower binding efficiencies, whilst GcvB-*ilvI* and GcvB-*ilvC* has consistently produced higher binding efficiencies.

Further, to show whether these interactions were specific, negative controls were added to the RNA array. These including spotting of the IVT DNA template for GcvB-MG on its own and inclusion of scrambled version of the target *ilvC* (*ilvC*-S), where the mRNA binding region was scrambled to disrupt GcvB interaction. The results of the control tests showed that GcvB-MG was only present where target mRNAs were immobilised, so no non-specific binding was evident, and by altering the GcvB-binding region of *ilvC*, to create *ilvC*-S, there was a drop in the fluorescence intensity of bound GcvB-MG of 90%, indicating binding of GcvB to *ilvC* to be specific.

With proof of concept demonstrated for using the RNA array technology to verify *in silico* predicted GcvB-mRNA target interactions, the next step for this work would be to use the approach to test the interactions of the other novel APP sRNAs identified in Chapter 1 with their potential mRNA partners, once predicted. Further, it's clear that coupling the bioinformatics steps with the RNA array technology generates a unique pipeline capable of both predicting and experimentally verifying sRNA-mRNA interactions. This has broad applications within the bacterial sRNA field, as well as the wider non-coding RNA research domain, where predicting and testing RNA-RNA interactions is crucial for unravelling molecular mechanism details.

6.4. Disruption of GcvB-mRNA interactions using a PNA

GcvB is recognised as an sRNA important in controlling metabolic pathways linked to virulence and pathogenesis. Consequently, anti-virulence/antibacterial strategies for treating bacterial infections, such as APP, could involve targeting GcvB to disrupt its interactions with mRNA targets to mediate a therapeutic effect. Therefore, having demonstrated the capability of the RNA array technology in validating the predicted GcvB-

mRNA target interactions, this work sought to design and test a PNA capable of specifically disrupting these interactions, monitoring for any PNA-disruption effects using the RNA array. The output of this work was both a PNA, designed to specifically target the G/U-rich seed region of GcvB, which specifically disrupted GcvB-mRNA interactions, as well as a proof-of-concept demonstration of using the RNA array for a sRNA-mRNA disruption assay.

The next steps for this work would be to test the GcvB-binding PNA *in vivo* in APP and explore whether it mediates an anti-virulence/antibacterial effect. However, given the concentrations required on the array, which may not translate to suitable levels for *in vivo* testing, it may be required to further improve the PNA. This could involve an iterative design process to support optimisation of the PNA binding to GcvB and disrupting its interactions with mRNAs, for example through altering nucleotide sequence composition and molecule length. Given the capability of the RNA array for testing interaction disruption, this approach could be used to rapidly test different PNA designs for impact on a range of GcvB-mRNA pairs.

Furthermore, applications testing off-target impact of PNA designs on broader RNA-RNA interactions, that may wish to be avoided, could also be undertaken. This could help determine the specificity of the PNA's effect on RNA-interactions and inform further re-design steps to improve the PNA's utility as a potential therapeutic. Finally, considering PNA uptake into cells, a range of peptide tags may need to be explored to support this, and inclusion of a fluorescent tag on the PNA can facilitate monitoring this using microscopy as one of the first steps.

6.5. Broader context: An emerging therapeutic discovery pipeline targeting bacterial sRNAs

Bacterial pathogenesis is often regulated by sRNA-controlled pathways (Svensson & Sharma, 2016). GcvB is recognised as one of these sRNAs, and therefore represents a suitable therapeutic target. Other well characterised sRNAs identified as antibacterial targets include the Quorum Regulatory RNAs (Qrr sRNAs), as these link quorum sensing systems and pathogenesis (Rutherford & Bassler, 2012; Defoirdt, 2018). Indeed, with new

antibacterial targets desperately needed, as antibiotic resistance to current drugs continues to rise (World Health Organisation, 2015), coupled with pace gaining in the RNA therapeutics area, bacterial RNAs, such as sRNAs involved in controlling critical pathways or virulence mechanisms, represent an untapped pool of potential targets for exploitation.

Excitingly, the approach of using antisense nucleic acid mimics (NAMs) to inhibit RNA targets has expanded in recent years. Whilst early attempts to use this approach suffered from issues with cellular uptake and NAM stability, advances have seen improvements in nuclease resistance, enhanced interaction capabilities, and improved delivery approaches using peptide tags to support uptake (Khvorova & Watts, 2017; Hegarty & Stewart, 2018).

Nevertheless, with the expansion of drug discovery moving into the RNA space, new approaches are required to discover sRNA antibacterial targets, as well as new methods to support high-throughput screening and testing work as part of understanding drug interactions with their target sRNAs, and impact on the sRNAs mechanism of action, such as impact on binding to mRNA partners. The bioinformatics pipeline developed in this thesis, coupled with the validation and PNA testing studies undertaken using the RNA array technology, represent suitable tools to fill the gap. Collectively, this work can be recognised as taking the first steps towards developing a therapeutic discovery pipeline targeting bacterial sRNAs (Figure 6.1).

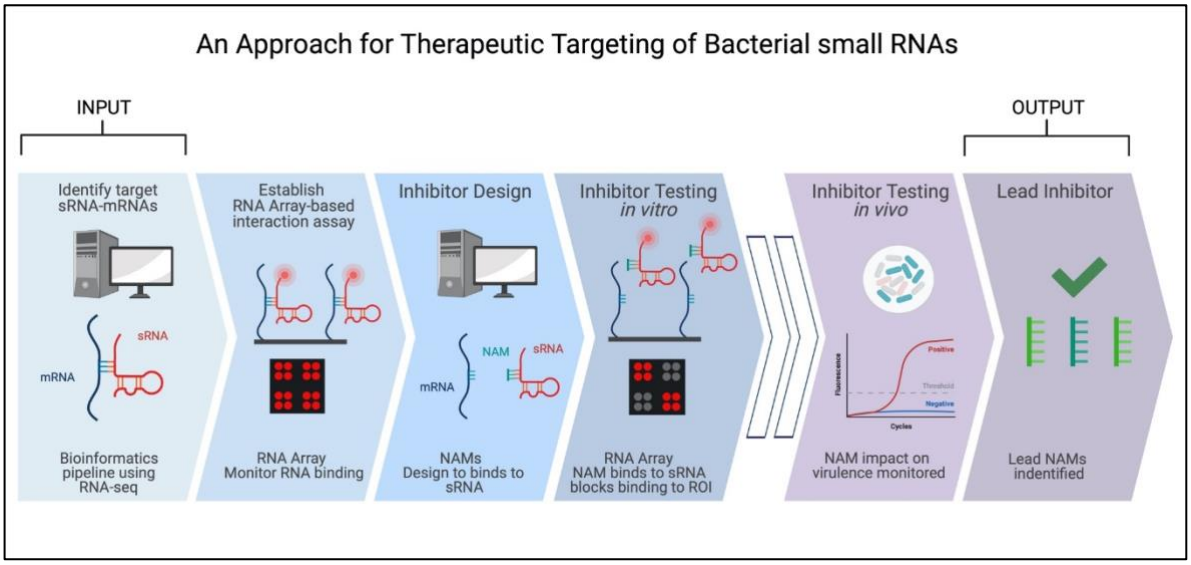


Figure 6.1. Schematic illustration of a therapeutic discovery pipeline targeting bacterial sRNAs. The input for the pipeline is RNA-seq data relevant to infectious/non-infectious samples. This is mined and assessed using various bioinformatic tools to predict a target sRNA, and associated mRNA partners. An understanding of the mRNA partners, coupled with RNA-seq conditions, can infer the relevance of the sRNA as an antibacterial target. Using the RNA array technology, an interaction assay is established to monitor the predicted sRNA-mRNA interactions. *In silico* approaches, coupled with bioinformatic hybridisation data and array interaction efficiency data for the sRNA-mRNA pairs, can help inform NAM design. The RNA array can be used to monitor sRNA-mRNA disruption following PNA addition. These steps were all completed as part of this study based on GcvB and its mRNA targets. The pipeline is completed by undertaking *in vivo* testing which results in a lead NAM being identified as an sRNA inhibitor. The pipeline may involve iterative repeats of the NAM design and testing steps in order to improve NAM design. Ultimately, the pipeline outputs can be taken forward in the development of antibacterial/therapeutic strategies.

References

Agriculture and Horticulture Development Board AHDB (2019). “UK pig facts and figures”. Retrieved from https://pork.ahdb.org.uk/prices-stats/published-reports/

Aertsen, A., & Michiels, C. W. (2004). Stress and How Bacteria Cope with Death and Survival. C*rit Rev Microbiol*, *30*(4), 263–273.

Altuvia, S., Zhang, A., Argaman, L., Tiwari, A., & Storz, G. (1998). The *Escherichia coli* OxyS regulatory RNA represses fhlA translation by blocking ribosome binding. *The EMBO Journal*, *17*(20), 6069–6075.

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G., Margalit, H., & Altuvia, S. (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Current Biology*, *11*(12), 941–950.

Auer, P. L., & Doerge, R. W. (2010). Statistical design and analysis of RNA sequencing data. *Genetics*, *185*(2), 405–416.

Bai, H., Xue, X., Hou, Z., Zhou, Y., Meng, J., & Luo, X. (2010). Antisense antibiotics: a brief review of novel target discovery and delivery. *Current Drug Discovery Technologies*, *7*(2), 76–85.

Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y. I., Encarnación, S., & Collado-Vides, J. (2009). Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiology Reviews, 33*(1).

Bandyra, K. J., Said, N., Pfeiffer, V., Górna, M. W., Vogel, J., & Luisi, B. F. (2012). The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Molecular Cell*, *47*(6), 943–953.

Bashyam, M. D., & Hasnain, S. E. (2004). The extracytoplasmic function sigma factors: role in bacterial pathogenesis. *Infection, Genetics and Evolution*, *4*(4), 301–308.

Bertucci, A., Manicardi, A., Corradini, R. (2012). *Advanced Molecular Probes for Sequence-Specific DNA Recognition. Soft and Biological Matter, 89–124.*

Blackall, P. J., Klaasen, H. L., van den Bosch, H., Kuhnert, P., & Frey, J. (2002). Proposal of a new serovar of Actinobacillus pleuropneumoniae: serovar 15. *Veterinary Microbiology*, *84*(1-2), 47–52.

Bossé, J. T., & MacInnes, J. I. (1997). Genetic and biochemical analyses of *Actinobacillus pleuropneumoniae* urease. *Infection and Immunity*, *65*(11), 4389–4394.

Bossé, J. T., & MacInnes, J. I. (2000). Urease activity may contribute to the ability of *Actinobacillus pleuropneumoniae* to establish infection. *Canadian Journal of Veterinary*, *64*(3), 145–150.

Bossé, J. T., Gilmour, H. D., & MacInnes, J. I. (2001). Novel genes affecting urease activity in *Actinobacillus pleuropneumoniae*. *Journal of Bacteriology*, *183*(4), 1242–1247.

Bossé, J. T., Janson, H., Sheehan, B. J., Beddek, A. J., Rycroft, A. N., Simon Kroll, J., & Langford, P. R. (2002). *Actinobacillus pleuropneumoniae*: pathobiology and pathogenesis of infection. *Microbes and Infection*, *4*(2), 225–235.

Bossé, J. T., Sinha, S., Li, M. S., O'Dwyer, C. A., Nash, J. H. E., Rycroft, A. N., et al. (2010). Regulation of pga Operon Expression and Biofilm Formation in *Actinobacillus pleuropneumoniae* by E and H-NS. *Journal of Bacteriology*, *192*(9), 2414–2423.

Bossé, J. T., Soares-Bazzolli, D. M., Li, Y., Wren, B. W., Tucker, A. W., Maskell, D. J., et al. (2014). The generation of successive unmarked mutations and chromosomal insertion of heterologous genes in *Actinobacillus pleuropneumoniae* using natural transformation. *PloS One*, *9*(11), e111252.

Bossé, J. T., Chaudhuri, R. R., Li, Y., Leanse, L. G., Fernandez Crespo, R., Coupland, P., *et al*. (2016). Complete Genome Sequence of MIDG2331, a Genetically Tractable Serovar 8 Clinical Isolate of *Actinobacillus pleuropneumoniae*. *Genome Announcements*, *4*(1).

Bossé, J. T., Li, Y., Sárközi, R., Gottschalk, M., Angen, Ø., Nedbalcova, K., et al. (2017). A Unique Capsule Locus in the Newly Designated *Actinobacillus pleuropneumoniae* Serovar 16 and Development of a Diagnostic PCR Assay. *Journal of Clinical Microbiology*, *55*(3), 902–907.

Bossé, J. T., Li, Y., Crespo, R. F., Angen, Ø., Holden, M. T. G., Weinert, L. A., et al. (2020). Draft Genome Sequences of the Type Strains of *Actinobacillus indolicus* (46K2C) and *Actinobacillus porcinus* (NM319), Two NAD-Dependent Bacterial Species Found in the Respiratory Tract of Pigs. *Microbiology Resource Announcements*, *9*(1).

Bouché, F., & Bouché, J. P. (1989). Genetic evidence that DicF, a second division inhibitor encoded by the *Escherichia coli* dicB operon, is probably RNA. *Molecular Microbiology*, *3*(7), 991–994.

Browning, D. F., & Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, *2*(1), 57–65.

Burgess, R. R., Travers, A. A., Dunn, J. J., & Bautz, E. K. F. (1969). Factor Stimulating Transcription by RNA Polymerase. *Nature*, *221*(5175), 43–46.

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., & McQuillan, J. A. (2012). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, *28*(4), 464–469.

Cases, I., & de Lorenzo, V. (2005). Promoters in the environment: transcriptional regulation in its natural context. *Nature Reviews Microbiology*, *3*(2), 105–118.

Chien, M.-S., Chan, Y.-Y., Chen, Z.-W., Wu, C.-M., Liao, J.-W., Chen, T.-H., et al. (2009). *Actinobacillus pleuropneumoniae* serotype 10 derived ApxI induces apoptosis in porcine alveolar macrophages. *Veterinary Microbiology*, *135*(3-4), 327–333.

Chiers, K., De Waele, T., Pasmans, F., Ducatelle, R., & Haesebrouck, F. (2010). Virulence factors of *Actinobacillus pleuropneumoniae* involved in colonization, persistence and induction of lesions in its porcine host. *Veterinary Research*, *41*(5), 65.

Cho, J. H., & Kim, I. H. (2011). Effect of stocking density on pig production. *African Journal of Biotechnology*, *10*(63), 13688–13692.

Chung, J. W., Ng-Thow-Hing, C., Budman, L. I., Gibbs, B. F., Nash, J. H. E., Jacques, M., & Coulton, J. W. (2007). Outer membrane proteome of *Actinobacillus pleuropneumoniae*: LC-MS/MS analyses validate *in silico* predictions. *Proteomics*, *7*(11), 1854–1865.

Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2016). GenBank. *Nucleic Acids Research*, *44*(D1), D67–D72.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Gen. Biology*, *17*(1), 1.

Crick, F. (1958). On protein synthesis. *Symp Soc Exp Biol*, 12: 38-163.

Crick, F. (1966). Codon-anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology*, 19(2), 548–555.

de Jong, A., Pietersma, H., Cordes, M., Kuipers, O. P., & Kok, J. (2012). PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics*, *13*(1), 299.

De Lay, N., & Gottesman, S. (2011). Role of polynucleotide phosphorylase in sRNA function in *Escherichia coli*. *RNA*, *17*(6), 1172–1189.

Defoirdt, T. (2018). Quorum-Sensing Systems as Targets for Antivirulence Therapy. *Trends in Microbiology*, *26*(4), 313–328.

Dersch, P., Khan, M. A., Mühlen, S., & Görke, B. (2017). Roles of Regulatory RNAs for Antibiotic Resistance in Bacteria and Their Potential Value as Novel Drug Targets. *Frontiers in Microbiology*, *8*, 803.

Deutscher, M. P. (2006). Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Research*, *34*(2), 659–666.

Deutscher, M. P. (2015). How bacterial cells keep ribonucleases under control. *FEMS Microbiology Reviews*, *39*(3), 350–361.

Diarra, M. S., Lavoie, M. C., Jacques, M., Darwish, I., Dolence, E. K., Dolence, J. A., et al. (1996). Species selectivity of new siderophore-drug conjugates that use specific iron uptake for entry into bacteria. *Antimicrobial Agents and Chemotherapy*, *40*(11), 2610–2617.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.

Feklístov, A., Sharon, B. D., Darst, S. A., & Gross, C. A. (2014). Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *Dx.Doi.org*, *68*(1), 357–376.

Felden, B., & Cattoir, V. (2018). Bacterial Adaptation to Antibiotics through Regulatory RNAs. *Antimicrobial agents and chemotherapy*, *62*(5), e02503-17.

Fender, A., Elf, J., Hampel, K., Zimmermann, B., & Wagner, E. G. H. (2010). RNAs actively cycle on the Sm-like protein Hfq. *Genes & Development*, *24*(23), 2621–2626.

Franze de Fernandez, M. T., Hayward, W. S., & August, J. T. (1972). Bacterial Proteins Required for Replication of Phage Qβ Ribonucleic Acid: PURIFICATION AND PROPERTIES OF HOST FACTOR I, A RIBONUCLEIC ACID-BINDING PROTEIN. *Journal of Biological Chemistry*, *247*(3), 824–831.

Frey, J. (1995). Virulence in *Actinobacillus pleuropneumoniae* and RTX toxins. *Trends in Microbiology*, *3*(7), 257–261.

Frey, J. (2003). Detection, identification, and subtyping of *Actinobacillus pleuropneumoniae*. *Methods in Molecular Biology (Clifton, N.J.)*, *216*, 87–95.

Fuller, T. E., Martin, S., Teel, J. F., Alaniz, G. R., Kennedy, M. J., & Lowery, D. E. (2000). Identification of *Actinobacillus pleuropneumoniae* virulence genes using signature-tagged mutagenesis in a swine infection model. *Microbial Pathogenesis*, *29*(1), 39–51.

Gao, X. (2003). Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Research*, *31*(22), 143e–143.

Geller, B. L., Deere, J., Antimicrobial, Tilley, L., Iversen, P. L. (2005). Antisense phosphorodiamidate morpholino oligomer inhibits viability of *Escherichia coli* in pure culture and in mouse peritonitis. *Journal of Antimicrobial Chemotherapy 55*(6), 983-988.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, *5*(10), R80–R80.

Georg, J., & Hess, W. R. (2011). cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiology and Molecular Biology Reviews*, *75*(2), 286–300.

Good, L., Awasthi, S. K., Dryselius, R., Larsson, O., & Nielsen, P. E. (2001). Bactericidal antisense effects of peptide–PNA conjugates. *Nature Biotechnology*, *19*(4), 360–364.

Gottesman, S. (2004). The Small RNA Regulators of *Escherichia coli*: Roles and Mechanisms. *Annu Rev of Microbiology, 58*, 303–328.

Gottesman, S., & Storz, G. (2011). Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harbor Perspectives in Biology*, *3*(12), a003798.

Gottschalk, M., Broes, A., Mittal, K. R., Kobisch, M., Kuhnert, P., Lebrun, A., & Frey, J. (2003). Non-pathogenic *Actinobacillus* isolates antigenically and biochemically similar to *Actinobacillus pleuropneumoniae*: a novel species? *Veterinary Microbiology*, *92*(1-2), 87–101.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, *38*(Web Server), W695–W699.

Gruber, A. R., Lorenz, R., Benhart, S. H., Neubock, R., (2008). Vienna RNA Websuite. *Nucleic Acids Research, 36*(Web Server), W70–W74.

Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J., et al. (2001). Binding of the Initiation Factor σ70 to Core RNA Polymerase Is a Multistep Process. *Molecular Cell*, *8*(1), 21–31.

Gruber, T. M., & Gross, C. A. (2003). Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu Rev Microbiol*, *57*(1), 441–466.

Grundy, F. J., Lehman, S. C., & Henkin, T. M. (2003). The L box regulon: Lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proceedings of the National Academy of Sciences*, *100*(21), 12057–12062.

Haesebrouck, F., Chiers, K., Van Overbeke, I., & Ducatelle, R. (1997). *Actinobacillus pleuropneumoniae* infections in pigs: the role of virulence factors in pathogenesis and protection. *Veterinary Microbiology*, *58*(2), 239–249.

Hausser, J., Mayo, A., Keren, L., & Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. *Nature Communications*, *10*(1), 1–15.

Hegarty, J. P., & Stewart, D. B. S. (2018). Advances in therapeutic bacterial antisense biotechnology. *Applied Microbiology and Biotechnology*, *102*(3), 1055–1065.

Henderson, C. A., Vincent, H. A., acids, C. S. N. (2013). Characterization of MicA interactions suggests a potential novel means of gene regulation by small non-coding RNAs. *Nucleic Acids Research, 41*(5), 3386–3397.

Henderson, C. A., Rail, C. A., Butt, L. E., Vincent, H. A., & Callaghan, A. J. (2019). Generation of small molecule-binding RNA arrays and their application to fluorogen-binding RNA aptamers. *Methods, 167*, 39–53.

Hensel, A., Stockhofe-Zurwieden, N., Petzoldt, K., & Lubitz, W. (1995). Oral immunization of pigs with viable or inactivated *Actinobacillus pleuropneumoniae* serotype 9 induces pulmonary and systemic antibodies and protects against homologous aerosol challenge. *Infect Immun*, *63*(8), 3048.

Hippel, von, P. H. (1998). An Integrated Model of the Transcription Complex in Elongation, Termination, and Editing. *Science*, *281*(5377), 660–665.

Holmqvist, E., Unoson, C., Reimegård, J., & Wagner, E. G. H. (2012). A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. *Molecular Microbiology*, *84*(3), 414–427.

Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews. RNA*, *8*(1), 10.1002/wrna.1364.

Hu, X.-P., Dourado, H., Schubert, P., & Lercher, M. J. (2020). The protein translation machinery is expressed for maximal efficiency in *Escherichia coli*. *Nature Communications*, *11*(1), 5260.

Imburgio, D., Rong, M., Ma, K., & McAllister, W. T. (2000). Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry*, *39*(34), 10419–10430.

Ishihama, A. (2010). Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiology Reviews*, *34*(5), 628–645.

Jacob, F., & Monod, J. (1961). On the Regulation of Gene Activity. *Cold Spring Harbor Symposia on Quantitative Biology*, *26*, 193–211.

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., et al. (2009). STRING 8 - a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, *37*(suppl_1), D412–D416.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*(Web Server issue), W5–9.

Kafri, M., Metzl-Raz, E., Jona, G., & Barkai, N. (2016). The Cost of Protein Production. *Cell Reports*, *14*(1), 22–31.

Kaminishi, T., Wilson, D. N., Takemoto, C., Harms, J. M., Kawazoe, M., Schluenzen, F., et al. (2007). A snapshot of the 30S ribosomal subunit capturing mRNA via the Shine-Dalgarno interaction. *Structure*, *15*(3), 289–297.

Kery, M. B., Feldman, M., Livny, J., & Tjaden, B. (2014). TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Research*, *42*(W1), W124–W129.

Khvorova, A., & Watts, J. K. (2017). The chemical evolution of oligonucleotide therapies of clinical utility. *Nature Biotechnology*, *35*(3), 238–248.

Kim, J. N., & Breaker, R. R. (2008). Purine sensing by riboswitches. *Biology of the Cell*, *100*(1), 1–11.

Kingsford, C. L., Ayanbule, K., & Salzberg, S. L. (2007). Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biology*, *8*(2), R22.

Klitgaard, K., Friis, C., Angen, Ø., & Boye, M. (2010). Comparative profiling of the transcriptional response to iron restriction in six serotypes of *Actinobacillus pleuropneumoniae* with different virulence potential. *BMC Genomics*, *11*(1), 698.

Konieczna, I., Zarnowiec, P., Kwinkowski, M., Kolesinska, B., Fraczyk, J., Kaminski, Z., & Kaca, W. (2012). Bacterial urease and its role in long-lasting human diseases. *Current Protein & Peptide Science*, *13*(8), 789–806.

Kröger, C., Colgan, A., Srikumar, S., Händler, K., Sivasankaran, S. K., Hammarlöf, D. L., et al. (2013). An Infection-Relevant Transcriptomic Compendium for *Salmonella* enterica Serovar Typhimurium. *Cell Host & Microbe*, *14*(6), 683–695.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645.

Lawal, A., Jejelowo, O., Chopra, A. K., & Rosenzweig, J. A. (2011). Ribonucleases and bacterial virulence. *Microbial Biotechnology*, *4*(5), 558–571.

Le Rhun, A., Beer, Y. Y., Reimegård, J., Chylinski, K., & Charpentier, E. (2016). RNA sequencing uncovers antisense RNAs and novel small RNAs in *Streptococcus pyogenes*. *RNA Biology*, *13*(2), 177–195.

Lease, R. A., & Belfort, M. (2000). A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. *Proceedings of the National Academy of Sciences*, *97*(18), 9919–9924.

Lewis, D. E. A., & Adhya, S. (2004). Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Molecular Microbiology*, *54*(3), 692–701.

Li, W., Ying, X., Lu, Q., & Chen, L. (2012). Predicting sRNAs and Their Targets in Bacteria. *Genomics, Proteomics & Bioinformatics*, *10*(5), 276–284.

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.

Lietard, J., Schaudy, E., Hölz, K., Ameur, D., & Somoza, M. M. (2019). High-Density DNA and RNA microarrays - Photolithographic Synthesis, Hybridization and Preparation of Large Nucleic Acid Libraries. *Journal of Visualized Experiments: JoVE*, (150).

Livny, J. (2012). Bioinformatic Discovery of Bacterial Regulatory RNAs Using SIPHT. *Methods in Molecular Biology (Clifton, N.J.), 905,* 3–14.

Lone, A. G., Deslandes, V., Nash, J. H. E., Jacques, M., & MacInnes, J. I. (2009). Modulation of Gene Expression in *Actinobacillus pleuropneumoniae* Exposed to Bronchoalveolar Fluid. *PloS One*, *4*(7), e6139.

Majdalani, N., Hernandez, D., & Gottesman, S. (2002). Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Molecular Microbiology*, *46*(3), 813–826.

Man, S., Cheng, R., Miao, C., Gong, Q., Gu, Y., Lu, X., et al. (2011). Artificial trans -encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Research*, *39*(8), e50–e50.

Mandin, P., & Guille, M. (2013). Expanding control in bacteria: interplay between small RNAs and transcriptional regulators to control gene expression. *Current Opinion in Microbiology*, *16*(2), 125–132.

Mann, M., Wright, P. R., & Backofen, R. (2017). IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Research, 45*(W1), W435–W439.

Marsteller, T. A., & Fenwick, B. (1999). *Actinobacillus pleuropneumoniae* disease and serology. *Journal of Swine Health and Production*, *7*(4), 161–165.

Massé, E., & Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *99*(7), 4620–4625.

Massé, E., Escorcia, F. E., & Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes & Development*, *17*(19), 2374–2383.

Matos, R. G., Casinhas, J., Bária, C., Santos, dos, R. F., Silva, I. J., & Arraiano, C. M. (2017). The Role of Ribonucleases and sRNAs in the Virulence of Foodborne Pathogens. *Frontiers in Microbiology*, *8*, 910.

Matos, R., Bária, C., Pobre, V., Andrade, J., & Arraiano, C. (2012). Exoribonucleases as Modulators of Virulence in Pathogenic Bacteria. *Frontiers in Cellular and Infection Microbiology*, *2*, 65.

Matsumoto, Y., Shigesada, K., Hirano, M., & Imai, M. (1986). Autogenous regulation of the gene for transcription termination factor rho in *Escherichia coli*: localization and function of its attenuators. *Journal of Bacteriology*, *166*(3), 945–958.

Mauger, D. M., Siegfried, N. A., & Weeks, K. M. (2013). The genetic code as expressed through relationships between mRNA structure and protein function. *FEBS Letters*, *587*(8), 1180–1188.

Menzel, A., Beyerbach, M., Siewert, C., Gundlach, M., Hoeltig, D., Graage, R., et al. (2014). *Actinobacillus pleuropneumoniae* challenge in swine: diagnostic of lung alterations by infrared thermography. *BMC Veterinary Research*, *10*(1), 199.

Miyakoshi, M., Chao, Y., & Vogel, J. (2015). Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *The EMBO Journal*, *34*(11), e201490546–1492.

Mizuno, T., Chou, M. Y., & Inouye, M. (1984). A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proceedings of the National Academy of Sciences*, *81*(7), 1966–1970.

Moller, T. (2002a). Hfq: A Bacterial Sm-like Protein that Mediates RNA-RNA Interaction. *Molecular Cell*, *9*(1), 23–30.

Moller, T. (2002b). Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes & Development*, *16*(13), 1696–1706.

Mooney, R. A., Darst, S. A., & Landick, R. (2005). Sigma and RNA Polymerase: An On-Again, Off-Again Relationship? *Molecular Cell*, *20*(3), 335–345.

Murakami, K. S., & Darst, S. A. (2003). Bacterial RNA polymerases: the whole story. *Current Opinion in Structural Biology*, *13*(1), 31–39.

Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., et al. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*, *43*(D1), D130–D137.

Norouzi, M., Pickford, A. R., Butt, L. E., Vincent, H. A., & Callaghan, A. J. (2019). Application of mRNA Arrays for the Production of mCherry Reporter-Protein Arrays for Quantitative Gene Expression Analysis. *ACS Synthetic Biology*, *8*(2), 207–215.

O'Neill, C., Jones, S. C. P., Bossé, J. T., Watson, C. M., Williamson, S. M., Rycroft, A. N., et al. (2010). Prevalence of *Actinobacillus pleuropneumoniae* serovars in England and Wales. *Veterinary Record*, *167*(17), 661–662.

Opriessnig, T., Giménez-Lirola, L. G., & Halbur, P. G. (2011). Polymicrobial respiratory disease in pigs. *Animal Health Research Reviews*, *12*(02), 133–148.

Phillips, J. O., Butt, L. E., Henderson, C. A., Devonshire, M., Healy, J., Conway, S. J., Locker, N., Pickford, A. R., Vincent, H. A., & Callaghan, A. J. (2018). High-density functional-RNA arrays as a versatile platform for studying RNA-based interactions. *Nucleic Acids Research, 46*(14), e86.

Pichon, C., & Felden, B. (2008). Small RNA gene identification and mRNA target predictions in bacteria. *Bioinformatics*, *24*(24), 2807–2813.

Pontes, M. H., Sevostyanova, A., & Groisman, E. A. (2015). When Too Much ATP Is Bad for Protein Synthesis. *Journal of Molecular Biology*, *427*(16), 2586–2594.

Proft, T., & Baker, E. N. (2009). Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cellular and Molecular Life Sciences*, *66*(4), 613–635.

Pulvermacher, S. C., Stauffer, L. T., Stauffer, G. V. (2008). The role of the small regulatory RNA GcvB in GcvB/mRNA posttranscriptional regulation of *oppA* and *dppA* in *Escherichia coli*. *FEMS Microbiology Letters*, 281(1), 42–50.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rajkumar, A. P., Qvist, P., Lazarus, R., Lescai, F., Ju, J., Nyegaard, M., et al. (2015). Experimental validation of methods for differential gene expression analysis and sample pooling in RNA-seq. *BMC Genomics*, *16*(1), 548.

Ramjeet, M., Deslandes, V., Gouré, J., & Jacques, M. (2008). *Actinobacillus pleuropneumoniae* vaccines: from bacterins to new insights into vaccination strategies. *Animal Health Research Reviews*, *9*(01), 25–45.

Rasmussen, L. C. V., Sperling-Petersen, H. U., & Mortensen, K. K. (2007). Hitting bacteria at the heart of the central dogma: sequence-specific inhibition. *Microbial Cell Factories*, *6*(1), 24.

Rayamajhi, N., Shin, S. J., Kang, S. G., Lee, D. Y., Ahn, J. M., & Yoo, H. S. (2016). Development and Use of a Multiplex Polymerase Chain Reaction Assay Based on Apx Toxin Genes for Genotyping of *Actinobacillus Pleuropneumoniae* Isolates. *Journal of Veterinary Diagnostic Investigation*, *17*(4), 359–362.

Repoila, F., & Darfeuille, F. (2009). Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biology of the Cell*, *101*(2), 117–131.

Rice, L. B. (2008). Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. *The Journal of Infectious Diseases*, 197(8), 1079–1081,

Rivas, E., Klein, R. J., Jones, T. A., & Eddy, S. R. (2001). Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Current Biology*, *11*(17), 1369–1373.

Roberts, J. W. (2019). Mechanisms of Bacterial Transcription Termination. *Journal of Molecular Biology*, *431*(20), 4030–4039.

Rodionov, D. A., Vitreschak, A. G., A. M. N. (2003). Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Research, 31*(23), 6748-57.

Rossi, C. C., Bossé, J. T., Li, Y., Witney, A. A., Gould, K. A., Langford, P. R., & Bazzolli, D. M. S. (2016). A computational strategy for the search of regulatory small RNAs in *Actinobacillus pleuropneumoniae*. *RNA*, *22*(9), 1373–1385.

Rudkin, J. K., McLoughlin, R. M., Preston, A., & Massey, R. C. (2017). Bacterial toxins: Offensive, defensive, or something else altogether? *PLOS Pathogens*, *13*(9), e1006452.

Rutherford, S. T., & Bassler, B. L. (2012). Bacterial *quorum* sensing: its role in virulence and possibilities for its control. *Cold Spring Harbor Perspectives in Medicine*, *2*(11).

Saecker, R. M., Record, M. T., Jr, & deHaseth, P. L. (2011). Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis. *Journal of Molecular Biology*, *412*(5), 754–771.

Sanches, N. M. (2018). *Análise in silico e caracterização funcional de RNAs pequenos reguladores: alvos e fenótipos envolvidos em Actinobacillus pleuropneumoniae.* (Doctoral dissertation). Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., et al. (2019). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *47*(D1), D23–D28.

Sárközi, R., Makrai, L., & Fodor, L. (2015). Identification of a proposed new serovar of *Actinobacillus Pleuropneumoniae*: Serovar 16. *Acta Veterinaria Hungarica*, *63*(4), 444–450.

Sharma, A. K., Dhasmana, N., Dubey, N., Kumar, N., Gangwal, A., Gupta, M., & Singh, Y. (2017). Bacterial Virulence Factors: Secreted for Survival. *Indian Journal of Microbiology*, *57*(1), 1–10.

Sharma, C. M., Darfeuille, F., Plantinga, T. H., & Vogel, J. (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes & Development*, *21*(21), 2804–2817.

Sharma, C. M., Papenfort, K., Pernitzsch, S. R., Mollenkopf, H. J., Hinton, J. C. D., & Vogel, J. (2011). Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Molecular Microbiology*, *81*(5), 1144–1165.

Sheehan, B. J., Bossé, J. T., Beddek, A. J., Rycroft, A. N., Kroll, J. S., & Langford, P. R. (2003). Identification of *Actinobacillus pleuropneumoniae*; Genes Important for Survival during Infection in Its Natural Host. *Infection and Immunity*, *71*(7), 3960.

Shine, J., & Dalgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. *Nature*, *254*(5495), 34–38.

Shope, R. E. (1964). Porcine contagious pleuropneumonia I experimental transmission, etiology, and pathology. *Journal of Experimental Medicine*, *119*(3), 357–368.

Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J., & Holmes, I. H. (2009). JBrowse: a next-generation genome browser. *Genome Research*, *19*(9), 1630–1638.

Snel, B., Lehmann, G., Bork, P., & Huynen, M. A. (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research*, *28*(18), 3442–3444.

Stauffer, L. T., & Stauffer, G. V. (2005). GcvA interacts with both the alpha and sigma subunits of RNA polymerase to activate the *Escherichia coli* gcvB gene and the gcvTHP operon. *FEMS Microbiology Letters, 242*(2), 333–338.

Sthitmatee, N., Sirinarumitr, T., Makonkewkeyoon, L., Sakpuaram, T., & Tesaprateep, T. (2003). Identification of the *Actinobacillus pleuropneumoniae* serotype using PCR based-apx genes. *Molecular and Cellular Probes*, *17*(6), 301–305.

Stonington, O. G., & Pettijohn, D. E. (1971). The folded genome of *Escherichia coli* isolated in a protein-DNA-RNA complex. *Proceedings of the National Academy of Sciences*, *68*(1), 6–9.

Storz, G. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell*, *43*(6), 880–891.

Storz, G., Opdyke, J. A., & Zhang, A. (2004). Controlling mRNA stability and translation with small, noncoding RNAs. *Current Opinion in Microbiology*, *7*(2), 140–144.

Stouthamer, A. H. (1973). A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie Van Leeuwenhoek*, *39*(3), 545–565.

Subashchandrabose, S., Leveque, R. M., Kirkwood, R. N., Kiupel, M., & Mulks, M. H. (2013). The RNA chaperone Hfq promotes fitness of *Actinobacillus pleuropneumoniae* during porcine pleuropneumonia. *Infection and Immunity*, *81*(8), 2952–2961.

Subashchandrabose, S., Leveque, R. M., Wagner, T. K., Kirkwood, R. N., Kiupel, M., & Mulks, M. H. (2009). Branched-Chain Amino Acids Are Required for the Survival and Virulence of *Actinobacillus pleuropneumoniae* in Swine. *Infection and Immunity*, *77*(11), 4925–4933.

Sully, E. K., & Geller, B. L. (2016). Antisense antimicrobial therapeutics. *Current Opinion in Microbiology*, *33*, 47–55.

Summerton, J. E. (2006). Morpholinos and PNAs Compared. In *Peptide Nucleic Acids, Morpholinos and Related Antisense Biomolecules* (pp. 89–113). Springer, Boston, MA.

Svensson, S. L., & Sharma, C. M. (2016). Small RNAs in Bacterial Virulence and Communication. *Microbiology Spectrum*, *4*(3), 10.1128/microbiolspec.VMBF-0028-2015.

Taboada, B., Estrada, K., Ciria, R., & Merino, E. (2018). Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, *34*(23), 4118–4120.

Takele Assefa, A., Vandesompele, J., & Thas, O. (2020). On the utility of RNA sample pooling to optimize cost and statistical power in RNA sequencing experiments. *BMC Genomics*, *21*(1), 312.

Thomason, M. K., & Storz, G. (2010). Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annual Review of Geneticx, 2010 44*(1), 167-188.

Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D., & Gally, D. L. (2014). Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic *Escherichia coli*. *Molecular Cell*, *55*(2), 199–213.

Updegrove, T. B., Shabalina, S. A., & Storz, G. (2015). How do base-pairing small RNAs evolve? *FEMS Microbiology Reviews*, *39*(3), 379–391.

Urbanowski, M. L., Stauffer, L. T., & Stauffer, G. V. (2000). The gcvB gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Molecular Microbiology*, *37*(4), 856–868.

VanderWaal, K., & Deen, J. (2018). Global trends in infectious diseases of swine. *Proc Natl Acad Sci USA*, *115*(45), 11495.

Vanni, M., Merenda, M., Barigazzi, G., Garbarino, C., Luppi, A., Tognetti, R., & Intorre, L. (2012). Antimicrobial resistance of *Actinobacillus pleuropneumoniae* isolated from swine. *Veterinary Microbiology*, *156*(1-2), 172–177.

Verona, M. D., Verdolino, V., Palazzesi, F., & Corradini, R. (2017). Focus on PNA Flexibility and RNA Binding using Molecular Dynamics and Metadynamics. *Scientific Reports*, *7*(1), 1–11.

Vincent, H. A., Phillips, J. O., Henderson, C. A., Roberts, A. J., Stone, C. M., Mardle, C. E., et al. (2013). An Improved Method for Surface Immobilisation of RNA: Application to Small Non-Coding RNA - mRNA Pairing. *PloS One*, *8*(11), e79142.

Vitreschak, A. G., Lyubetskaya, E. V., Shirshin, M. A., Gelfand, M. S., & Lyubetsky, V. A. (2004). Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiology Letters, 234*(2), 357–370.

Vogel, J. (2009). A rough guide to the non-coding RNA world of *Salmonella*. *Molecular Microbiology*, *71*(1), 1–11.

Vogel, J., & Luisi, B. F. (2011). Hfq and its constellation of RNA. *Nature Reviews Microbiology*, *9*(8), 578–589.

Vogel, J., & Papenfort, K. (2006). Small non-coding RNAs and the bacterial outer membrane. *Current Opinion in Microbiology*, *9*(6), 605–611.

Vogel, J., & Sharma, C. M. (2005). How to find small non-coding RNAs in bacteria. *Biol Chem*, *386*(12), 1219–1238.

Vogel, J., & Wagner, E. G. H. (2007). Target identification of small noncoding RNAs in bacteria. *Current Opinion in Microbiology*, *10*(3), 262–270.

Ward, C. K., & Inzana, T. J. (1994). Resistance of *Actinobacillus pleuropneumoniae* to bactericidal antibody and complement is mediated by capsular polysaccharide and blocking antibody specific for lipopolysaccharide. *Journal of Immunology*, *153*(5), 2110–2121.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.

Waters, L. S., & Storz, G. (2009). Regulatory RNAs in Bacteria. *Cell*, *136*(4), 615–628.

Westermann, A. J., Venturini, E., Sellin, M. E., Förstner, K. U., Hardt, W.-D., & Vogel, J. (2019). The Major RNA-Binding Protein ProQ Impacts Virulence Gene Expression in *Salmonella enterica* Serovar *Typhimurium*. *mBio*, *10*(1), e02504-1.

Wilson, K. S., & Hippel, von, P. H. (1995). Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc Natl Acad Sci USA*, *92*(19), 8793.

Winkler, W. C., & Breaker, R. R. (2003). Genetic Control by Metabolite-Binding Riboswitches. *ChemBioChem*, *4*(10), 1024–1032.

World Health Organisation (2015). "Global action plan on antimicrobial resistance". Retrieved from https://www.who.int/antimicrobial-resistance/publications/global-action-plan/en/

Wojciechowska, M., Równicki, M., Mieczkowski, A., Miszkiewicz, J., & Trylska, J. (2020). Antibacterial Peptide Nucleic Acids-Facts and Perspectives. *Molecules, 25*(3), 559.

Xu, Z., Chen, X., Li, L., Li, T., Wang, S., Chen, H., & Zhou, R. (2010). Comparative genomic characterization of *Actinobacillus pleuropneumoniae*. *Journal of Bacteriology*, *192*(21), 5625–5636.

Zampetaki, A., Albrecht, A., & Steinhofel, K. (2018). Long Non-coding RNA Structure and Function: Is There a Link? *Frontiers in Physiology*, *9*, 121.

Zhang, B., Feng, S., Xu, C., Zhou, S., He, Y., Zhang, L., et al. (2012). Serum resistance in *Haemophilus parasuis* SC096 strain requires outer membrane protein P2 expression. *FEMS Microbiology Letters*, *326*(2), 109–115.

Zundel, M. A., Basturea, G. N., & Deutscher, M. P. (2009). Initiation of ribosome degradation during starvation in *Escherichia coli*. *RNA*, *15*(5), 977–983.

Appendices

Appendix 1 – List of predicted sRNAs provided by Imperial College London (ICL_UoV_2016).

ID	Gene	Location	Size	Strand
ARRC01/RNA05	GcvB	149359..149554	202	complement
ARRC02		418351..418426	75	forward
ARRC03	Lys Riboswitch	522806..522977	171	forward
ARRC05	MOCO_RNA	760650..760819	169	complement
ARRC06	SRP small	896728..896872	144	forward
ARRC07	Gly Riboswitch	1279189..1279418	229	complement
ARRC08		1997375..1997567	192	forward
ARRC10	6S	131769..131951	182	complement
ARRC11		1997375..1997567	237	forward
ARRC12		249050..249119	69	forward
ARRC13/RNA11	FMN riboswitch	440504..440700	196	forward
ARRC14/RNA08		451950..452112	162	forward
ARRC15	RNaseP	563243..563676	433	forward
ARRC16		1710971..1711121	150	forward
ARRC16b		1711050..1711121	72	forward
ARRC17	Alpha_RBS	2064968..2065114	146	forward
ARRC19	His leader	2308407..2308602	195	forward
ARRC22	MOCO_RNA	762120..762308	188	forward
cspA	cspA	137954..138368	414	complement
Lys riboswitch	Lys riboswitch	791357..791578	221	forward
RNA01		738604..738689	85	complement
RNA02		662472..662552	80	forward
RNA03		2035731..2035807	76	forward
RNA04		358279..358387	108	forward
RNA07	S15	1144541..1144667	126	forward
RNA10		1996019..1996142	123	forward
RNA12		2292548..2292720	172	forward
RNA12a		2292439..2292513	74	forward
RNA13	RNA_out	1546752..1546840	88	complement
RNA14	RNA_out	1555757..1555844	87	forward
RNA15		123028..123186	158	forward
RNA16		195041..195194	153	complement
RNA17		516723..516916	193	forward
RNA18		562723..562809	86	forward

ID	Gene	Location	Size	Strand
RNA19		808647..808776	129	complement
RNA20		896366..896461	95	complement
RNA21		1011268..1011367	99	forward
RNA22		1231995..1232196	201	complement
RNA23		1345522..1345653	131	forward
RNA24		1482564..1482686	122	forward
RNA25		1558420..1558683	263	complement
RNA26		1558660..1558900	240	complement
RNA27		1565840..1565919	79	forward
RNA28		1847012..1847242	230	complement
RNA29		1980934..1981119	185	forward
RNA30		2021324..2021575	251	complement
RNA31		2129217..2129562	345	complement
SECIS_3	SECIS_3	1003073..1003138	65	forward
t44	t44	602581..602722	141	forward
TPP	TPP	195281..195375	94	forward
TPP	TPP	567060..567171	111	forward
JB_01		674005..674232	227	forward
JB_02		883531..883613	82	forward
JB_03		911548..911651	103	complement
JB_04		968916..969087	171	complement
JB_05		1432323..1432455	132	complement
JB_06		60804..60906	102	forward
JB_07		205917..206272	355	forward
JB_08		217434..217566	132	complement
JB_09		223800..223873	73	complement
JB_10		249020..249119	99	forward
JB_11		262647..262775	128	complement
JB_12		269386..269460	74	complement
JB_13		351320..351534	214	forward
JB_14		453832..453909	77	forward
JB_15		521540..521801	261	forward
JB_16		521839..522126	287	forward
JB_17		613410..613706	296	complement
JB_18		804971..805056	85	forward
JB_19		896438..896872	434	forward
JB_20		1002322..1002396	74	forward
JB_21		1130623..1130880	257	complement
JB_22		1130864..1131096	232	forward
JB_23		1225977..1226060	83	complement

ID	Gene	Location	Size	Strand
JB_24		1245037..1245264	227	forward
JB_25	C4	1435557..1435797	240	forward
JB_26		1514415..1514489	74	complement
JB_27		1654302..1654387	85	complement
JB_28		1821900..1822016	116	forward
JB_29		1826334..1826438	104	forward
JB_30		1836223..1836481	258	forward
JB_31		1843180..1843494	314	Forward
JB_32		1847162..1847293	131	Forward
JB_33		1947542..1947634	92	complement
JB_34		2003547..2003623	76	forward
JB_35		2026728..2026907	179	complement
JB_36		2035311..2035394	83	forward
JB_37		2053974..2054153	179	forward
tmRNA	tmRNA	2118112..2118483	371	Forward
JB_38		2129671..2129760	89	complement
JB_39		2218790..2219245	455	complement
JB_40		2263805..2263903	98	forward
JB_41		2286993..2287141	148	forward

Appendix 2 – Complete list of novel sRNAs generated in this thesis (UoP_ICL_UoV_2020).

1 - Due to their size and content, the data frame for ICL_UoP_2020 with full details and re-annotated reference sequence for MIDG2331 (LN908249.1) are stored in a GitHub Repository at:

https://github.com/GenDataPro/APP_sRNAs_2021

2 - Link for RNA-seq data from APP serovar 8 MIDG2331 mapped to the LN908249.1 reference sequence genome in JBrowse:

http://hinton-jbrowse.s3-website-eu-west-1.amazonaws.com/JBrowse/index.html?data=00_SAMPLES/app-2017/app/data

(generated and hosted by the Hinton Lar at the University of Liverpool)

Appendix 3 – NCBI accession numbers of Proteobacteria.

Serovar	Family	Ref. Seq.
Haemophilus parainfluenzae T3T1	*Pasteurellaceae*	NC_015964
Haemophilus parasuis SH0165	*Pasteurellaceae*	NC_011852
Haemophilus influenzae R2846	*Pasteurellaceae*	CP002276
Actinobacillus pleuropneumoniae 8 MIDG2331	*Pasteurellaceae*	NZ_LN908249.1
Pasteurella multocida Pm70	*Pasteurellaceae*	NC_002663
Mannheimia haemolytica M42548	*Pasteurellaceae*	NC_021082
Actinobacillus suis H91-0380	*Pasteurellaceae*	NC_018690
Salmonella enterica serovar *Typhimurium* SL1344	*Enterobacteriaceae*	NC_016810.1
Escherichia coli str. K-12 MG1655	*Enterobacteriaceae*	MF521836.1
Proteus mirabilis HI4320	*Morganellaceae*	NR_075212.1
Yersinia pestis D182038	*Yersiniaceae*	NC_017160.1
Vibrio cholerae 12129	*Vibrionaceae*	NZ_ACFQ00000000.1

Appendix 4 – NCBI accession numbers of APP serovars.

Serovar	Ref. Seq.	Size (Mb)	GC%	Protein	Gene
APP serovar 1 str. 4074	NZ_AACK00000000.1	2.29	41.4	1987	2273
APP serovar 2 str. 4226	NZ_ADXN00000000.1	2.31	41.2	2100	2182
APP serovar 3 str. JL03	NC_010278.1	2.24	41.2	1972	2126
APP serovar 4 str. M62	NZ_ADOF00000000.1	2.27	41.2	2203	2253
APP serovar 5b str. L20	NC_009053.1	2.27	41.3	2030	2171
APP serovar 6 str. Femo	NZ_ADXO00000000.1	2.38	41	2168	2241
APP serovar 7 str. AP76	NC_010939.1	2.33	41.2	2079	2232
APP serovar 8 MIDG2331	NZ_LN908249.1	2.34	41.1	2106	2247
APP serovar 9 str. CVJ13261	NZ_ADOI00000000.1	2.26	41.2	2188	2244
APP serovar 10 str. D13039	NZ_ADOJ00000000.1	2.27	41.2	2158	2210
APP serovar 11 str. 56153	NZ_ADOK00000000.1	2.27	41.2	2179	2232
APP serovar 12 str. 1096	NZ_ADOL00000000.1	2.19	41.2	2082	2134
APP serovar 13 str. N273	NZ_ADOM00000000.1	2.25	41.2	2143	2198

Appendix 5 – MultiQC statistical report from RNA-seq data.

File Name	Growth Condition	Batch	% Aligned	% GC	Length	Reads (m)
Total_WT	BHI Plate Aerobic Total Wild Type	Vertis 1	99.90%	51%	104	1.1
Total_WT_anaerobic	BHI Plate Anaerobic Total Wild Type	Vertis 1	100.00%	51%	95	1.3
2331WT_S1_R1_001	BHI Plate Wild Type Aerobic	Vertis 2	89.70%	42%	75	22.6
2331WT_S1_R1_001-CTGAAGCT	BHI Plate Wild Type Aerobic	Vertis 2	99.70%	38%	55	3.7
2331WT_S1_R1_001-TAATGCGC	BHI Plate Wild Type Aerobic	Vertis 2	95.30%	37%	54	1.6
2331hfq_S2_R1_001	BHI Plate Aerobic Delta-Hfq	Vertis 2	94.10%	38%	75	19.7
2331hfq_S2_R1_001-CTGAAGCT	BHI Plate Aerobic Delta-Hfq	Vertis 2	99.10%	36%	54	3.9
2331hfq_S2_R1_001-TAATGCGC	BHI Plate Aerobic Delta-Hfq	Vertis 2	97.10%	36%	54	1.6
2331BR_S5_R1_001	BHI Broth Aerobic Wild Type	Vertis 2	95.00%	39%	75	24.3
2331BR_S5_R1_001-CTGAAGCT	BHI Broth Aerobic Wild Type	Vertis 2	99.70%	34%	55	2.5
2331BR_S5_R1_001-TAATGCGC	BHI Broth Aerobic Wild Type	Vertis 2	98.40%	34%	54	1.8
2331AMP_S6_R1_001	BHI Broth Aerobic Ampicillin	Vertis 2	94.60%	40%	75	24.3
2331AMP_S6_R1_001-CTGAAGCT	BHI Broth Aerobic Ampicillin	Vertis 2	99.70%	34%	55	2.2
2331AMP_S6_R1_001-TAATGCGC	BHI Broth Aerobic Ampicillin	Vertis 2	98.00%	34%	54	1.5
2331TY_S7_R1_001	BHI Broth Aerobic Tylosin	Vertis 2	95.50%	40%	75	19.4
2331TY_S7_R1_001-CTGAAGCT	BHI Broth Aerobic Tylosin	Vertis 2	99.70%	35%	55	1.7
2331TY_S7_R1_001-TAATGCGC	BHI Broth Aerobic Tylosin	Vertis 2	98.60%	34%	54	1.6
RNA-BHI_S7_R1_001	BHI Broth Aerobic Wild Type	Vertis 3	98.60%	39%	150	14.6
RNA-BHI_S7_R3_001	BHI Broth Aerobic Wild Type	Vertis 3	98.00%	39%	150	14.6
RNA-PS_S8_R1_001	Porcine Serum Aerobic Wild Type	Vertis 3	99.50%	41%	150	13.5
RNA-PS_S8_R3_001	Porcine Serum Aerobic Wild Type	Vertis 3	99.00%	41%	150	13.5

Appendix 6 – STRING enrichment for potential mRNA targets of the putative *trans*-acting sRNA GcvB MIDG2331.

Biological Processes					
GO term ID	Term Description	Matching Proteins in the Network	Observed Gene Count	Background Gene Count	Strength
GO:1901607	alpha-amino acid biosynthetic process	*ilvD, leuC, dapA, trpD, ilvC, dapE, hisG*	7	30	1.19
GO:0044283	small molecule biosynthetic process	*ilvD, leuC, dxs, dapA, trpD, ilvC, dapE, hisG*	8	82	0.82
GO:0009082	branched-chain amino acid biosynthetic process	*ilvD, leuC, ilvC*	3	4	1.7
GO:0044281	small molecule metabolic process	*ilvD, leuC, dxs, dapA, trpD, ureA, ilvC, dapE, hisG*	9	160	0.58
GO:1901564	organonitrogen compound metabolic process	*ilvD, leuC, dxs, dapA, gcp, trpD, ureA, ilvC, dapE, hisG*	10	213	0.5
GO:0006549	isoleucine metabolic process	*ilvD, ilvC*	2	2	1.83
GO:0009097	isoleucine biosynthetic process	*ilvD, ilvC*	2	2	1.83
GO:0009099	valine biosynthetic process	*ilvD, ilvC*	2	2	1.83
GO:0019877	diaminopimelate biosynthetic process	*dapA, dapE*	2	4	1.53
GO:0009089	lysine biosynthetic process via diaminopimelate	*dapA, dapE*	2	5	1.43
GO:1901566	organonitrogen compound biosynthetic process	*ilvD, leuC, dxs, dapA, trpD, ilvC, dapE, hisG*	8	181	0.47
Molecular Processes					
GO term ID	Term Description	Matching Proteins in the Network	Observed Gene Count	Background Gene Count	Strength
GO:0046872	metal ion binding	*ilvD, leuC, dxs, gcp, trpD, ureA, ilvC, dapE, hisG*	9	106	0.76
GO:0016836	hydro-lyase activity	*ilvD, leuC, dapA*	3	11	1.26
GO:0003824	catalytic activity	*ilvD, leuC, dxs, dapA, gcp, trpD, ureA, ilvC, dapE, hisG*	10	280	0.38
GO:0016811	hydrolase activity, acting on carbon-nitrogen bonds	*ureA, dapE*	2	6	1.35

Appendix 7 – List of potential mRNA targets of the putative *trans*-acting sRNA GcvB MIDG2331.

List Rank	Predicted Target	Description	Energy (kcal/mol)	*p*-value	Interaction Region in the mRNA	Similarity to MIDG2331
1	*arcD*	arginine/ornithine antiporter	-18.62	0	-29 to -14	100%
2	*ilvC*	ketol-acid reductoisomerase	-18.15	0	-62 to -44	100%
3	*ilvI*	acetolactate synthase 3 catalytic subunits	-15.74	0	-41 to -26	100%
4	*dapE*	succinyl-diaminopimelate desuccinylase	-14.6	0.001	+3 to 19	100%
5	*thrC*	threonine synthase	-13.03	0.004	+5 to 20	100%
6	*fabG*	3-ketoacyl-ACP reductase	-12.73	0.005	-67 to -53	100%
7	*ilvD*	dihydroxy-acid dehydratase	-12.55	0.005	-32 to -20	100%
8	*menA*	1,4-dihydroxy-2-naphthoate octaprenyltransferase	-12.1	0.007	-19 to -5	92%
9	***hisD***	histidinol dehydrogenase	-11.53	0.01	-9 to 9	100%
10	*serC*	phosphoserine aminotransferase	-11.42	0.011	-10 to +6	100%
11	*ftsK*	DNA translocase FtsK	-11.34	0.011	-48 to -37	100%
12	*dxs*	1-deoxy-D-xylulose-5-phosphate synthase	-11.04	0.013	-5 to +12	100%
13	*plpA*	outer membrane lipoprotein 1	-10.77	0.016	-32 to -15	100%
14	*ureA*	urease subunit gamma	-10.64	0.017	-28 to -15	100%
15	*sfsA*	sugar fermentation stimulation protein A	-10.55	0.018	-8 to +7	100%
16	***hisC***	histidinol-phosphate aminotransferase	-10.32	0.02	2 to 20	100%
17	*serB*	phosphoserine phosphatase	-9.95	0.024	-2 to +15	100%

List Rank	Predicted Target	Description	Energy (kcal/mol)	*p*-value	Interaction Region in the mRNA	Similarity to MIDG2331
18	*dapA*	dihydrodipicolinate synthase	-9.72	0.027	-21 to -2	100%
19	***hisI***	phosphoribosyl-ATP pyrophosphatase	-9.65	0.028	-9 to 12	100%
20	*gcp*	DNA-binding/iron metalloprotein/AP endonuclease	-9.49	0.03	-10 to +9	100%
21	*hisG*	ATP phosphoribosyltransferase	-9.21	0.034	-25 to -10	100%
22	*leuC*	isopropylmalate isomerase large subunit	-9.14	0.035	-46 to -26	100%
23	*gntR*	HTH-type transcriptional regulator	-9.02	0.037	-3 to +15	100%
24	*secE*	preprotein translocase subunit SecE	-8.87	0.039	-79 to -64	100%
25	*lrp*	leucine-responsive transcriptional regulator	-8.66	0.043	-76 to -63	100%
26	*ilvE*	branched-chain amino acid aminotransferase	-8.5	0.046	-12 to +8	100%
27	*uxaC*	uronate isomerase	-8.49	0.046	-43 to -30	100%
28	*serA*	D-3-phosphoglycerate dehydrogenase	-8.4	0.048	-24 to -10	100%
29	*tehB*	tellurite resistance protein TehB	-8.35	0.049	-8 to +10	100%
30	***trpD***	anthranilate phosphoribosyltransferase	-8.35	0.049	-69 to -54	100%

Appendix 8 – IntaRNA interaction location results for **M**odified **G**cvB (GcvB-**MG**) against mRNA targets.

Target	Start(T)	End(T)	Query	Start(Q)	End(Q)	Energy
ilvI	11	37	GcvB-MG	63	89	-23.60
ilvC	14	36	GcvB-MG	73	95	-22.88
thrC	41	62	GcvB-MG	72	91	-11.09
serC	31	40	GcvB-MG	80	89	-10.46
hisG	58	68	GcvB-MG	77	87	-9.23
serB	52	88	GcvB-MG	61	92	-8.46
ilvE	33	59	GcvB-MG	62	85	-8.29
ureA	10	21	GcvB-MG	78	88	-8.29
tehB	105	127	GcvB-MG	72	91	-7.59

Appendix 9 – DNA IVT templates and overlapping oligonucleotides.

Table Legend

Legend	Description
Blue	T7 promoter
Grey	Three Gs between promoter and sequence recommended for IVT
Orange	GcvB putative seed region (RNATarget 2)
Red	Predicted translation start site, normally ATG
Black	Actual sequence e.g. mRNA target sequence including 5' UTR
Green	Streptavidin aptamer (SA)
Purple	SA Middle Linker containing GT repetition
Yellow	SA End Linker
Underline	Oligonucleotides overlapping region
Underline	Overlapping region for 3' Alexa Fluorophore oligonucleotide
FW	Forward oligonucleotides
RV	Reverse oligonucleotides
MW	Molecular weight of the DNA IVT template
MW (B)	MW of the DNA IVT template with biotin
MW (F)	MW of the DNA IVT template with fluorophore and biotin
RNA MW	Molecular weight of the RNA product

Molecular weight for all molecules were calculated using OligoCalc web tool at:

(https://horizondiscovery.com/en/products/tools/TM-Calculator)

3' end standard oligonucleotides Used in all constructs for slide immobilisation on DNA slide		
Biotin T7 Promoter	Biotin FW Oligo	5' – Biotin - **cgaggcagatct**TAATACGACTCACTATAG – 3'
5' end standard oligonucleotides Used in all constructs for slide immobilisation on RNA slide and detection		
SA Middle Linker + SA Aptamer + SA End Linker		
Streptavidin Aptamer (SA)	Full SA Linker Design	ACACACACACACACACACGCATGCAT ACCGACCAGAATCATGCAAGTGCGTAAG ATAGTCGCGGGCCGGGATGCATGC
	SA Aptamer Reverse Oligo (SA RV)	GCATGCATCCCGGCCCGCGACTATCTTACG CACTTGCATGATTCTGGTCGGTATGCATGCG TG
Alexa Fluorophore	Alexa Reverse Oligo (Alexa)	5' – Alexa(488, 532 or 647) – GCATGCATCCCGGCCCGCGACTATCTTAC GCA – 3'

DNA IVT templates of mRNA targets			
RNA name	**Sequences**		
1. mRNA 5'UTR *ilvE*	DNA IVT template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGTAAAAATTCATTTGTAATACATAACAATAAGGACACAACAACATGGCTTTAAAAGATTTAGACTGGGCCAATTTAGGCTTCTCATATACACACACACACACACACACGCATGCATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCCGGGATGCATGC *ilvE* construct length = 199 MW(B): 123634 MW(F): 124830 GC = 43%	
	Primers:	**ilvE FW1** Length = 85	TAATACGACTCACTATAGGGTAAAAATTCATTTGTAATACATAACAATAAGGACACAACAACATGGCTTTAAAAGATTTAGACTG
		ilvE FW2 Length = 83	CATGGCTTTAAAAGATTTAGACTGGGCCAATTTAGGCTTCTCATATACACACACACACACACACACGCATGCATACCGACCAG
		SA RV	GCATGCATCCCGGCCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGTATGCATGCGTG
2. mRNA 5'UTR *ilvI*	DNA IVT template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGTCATTATAAAGCAAACACAATATTTCACTTAAAAATCGTTGGAGGATTCAATGAAAAAACTTTCCGGAGCAGAAATGGTTGTTCAATCCTTGAAAACACACACACACACACACACGCATGCATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCCGGGATGCATGC *ilvI* construct length = 207 MW(B): 128577 MW(F): 129773 GC = 43%	
	Primers:	**ilvI FW1** Length = 87	TAATACGACTCACTATAGGGTCATTATAAAGCAAACACAATATTTCACTTAAAAATCGTTGGAGGATTCAATGAAAAAACTTTCCGG
		ilvI FW2 Length = 88	GATTCAATGAAAAAACTTTCCGGAGCAGAAATGGTTGTTCAATCCTTGAAAACACACACACACACACACACGCATGCATACCGACCAG
		SA RV	GCATGCATCCCGGCCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGTATGCATGCGTG

3. mRNA 5'UTR *thrC*	DNA IVT template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGCAAATCACTCTTAATAC TTAATAATAAAGGACTTTTCCTAATGAACTTATACAACATCAAACAC CCAGAAGAACAAGTTAATTTCACACACACACACACACACACGCAT GCATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGG CCGGGATGCATGC *thrC* construct length = 197 MW(B): 122397 MW(F): 123593 GC = 42%	
	Primers:	**thrC FW1** Length = 82	TAATACGACTCACTATAGGGCAAATCACTCTTAATA CTTAATAATAAAGGACTTTTCCTAATGAACTTATAC AACATCAAAC
		thrC FW2 Length = 85	CTAATGAACTTATACAACATCAAACACCCAGAAGA ACAAGTTAATTTCACACACACACACACACACGC ATGCATACCGACCAG
		SA RV	GCATGCATCCCGGCCCGCGACTATCTTACGCACT TGCATGATTCTGGTCGGTATGCATGCGTG
4. mRNA 5'UTR *serB*	DNA template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGAAAGCACATTATCTATG CACCATTGAAAAAAGTCTATTAAAAAGAGAACATTATGCCAAACA CGATTTTTATCTTCGCTAAATCTCTGACAACACAAACACACACACA CACACACACGCATGCATACCGACCAGAATCATGCAAGTGCGTAA GATAGTCGCGGGCCGGGATGCATGC *serB* construct length = 209 MW(B): 129812 MW(F): 131008 GC = 43%	
	Primers:	**serB FW1** Length = 84	TAATACGACTCACTATAGGGAAAGCACATTATCTAT GCACCATTGAAAAAAGTCTATTAAAAAGAGAACA TTATGCCAAACACG
		serB FW2 Length = 91	GAGAACATTATGCCAAACACGATTTTTATCTTCGC TAAATCTCTGACAACACAAACACACACACACACAC ACACGCATGCATACCGACCAG
		SA RV	GCATGCATCCCGGCCCGCGACTATCTTACGCACT TGCATGATTCTGGTCGGTATGCATGCGTG

5. mRNA 5'UTR *serC*	DNA template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGCTAAATCAGTTTTTAAA AACACACAAAGAGGCAAACACAATGACACAAGTTTATAATTTCAG CGCAGGTCCGGCAATGATGCCGACACACACACACACACACACGC ATGCATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGG GCCGGGATGCATGC *serC* construct length = 196 MW(B): 121788 MW(F): 122985 GC = 47%	
	Primers:	**serC FW1** Length = 84	TAATACGACTCACTATAGGGCTAAATCAGTTTTTAA AAACACACAAAGAGGCAAACACAATGACACAAG TTTATAATTTCAGCG
		serC FW2 Length = 80	GACACAAGTTTATAATTTCAGCGCAGGTCCGGCA ATGATGCCGACACACACACACACACACGCATG CATACCGACCAG
		SA Aptamer	GCATGCATCCCGGCCCGCGACTATCTTACGCACT TGCATGATTCTGGTCGGTATGCATGCGTG
6. mRNA 5'UTR *tehB*	DNA template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGTTATCATCATTATTTTAT TCGTTAGAGAGGATTAATTATGCAAAACTTAATTTGCTATAAACAA ATGCCGGTTTGGACCAAAACACACACACACACACACGCATGC ATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCC GGGATGCATGC *tehB* construct length = 194 MW(B): 120544 MW(F): 121740 GC = 42%	
	Primers:	**tehB FW1** Length = 89	TAATACGACTCACTATAGGGTTATCATCATTATTTTA TTCGTTAGAGAGGATTAATTATGCAAAACTTAATT TGCTATAAACAAATGCC
		tehB FW1 Length = 74	CTTAATTTGCTATAAACAAATGCCGGTTTGGACCA AAACACACACACACACACACGCATGCATACCG ACCAG
		SA Aptamer	GCATGCATCCCGGCCCGCGACTATCTTACGCACT TGCATGATTCTGGTCGGTATGCATGCGTG

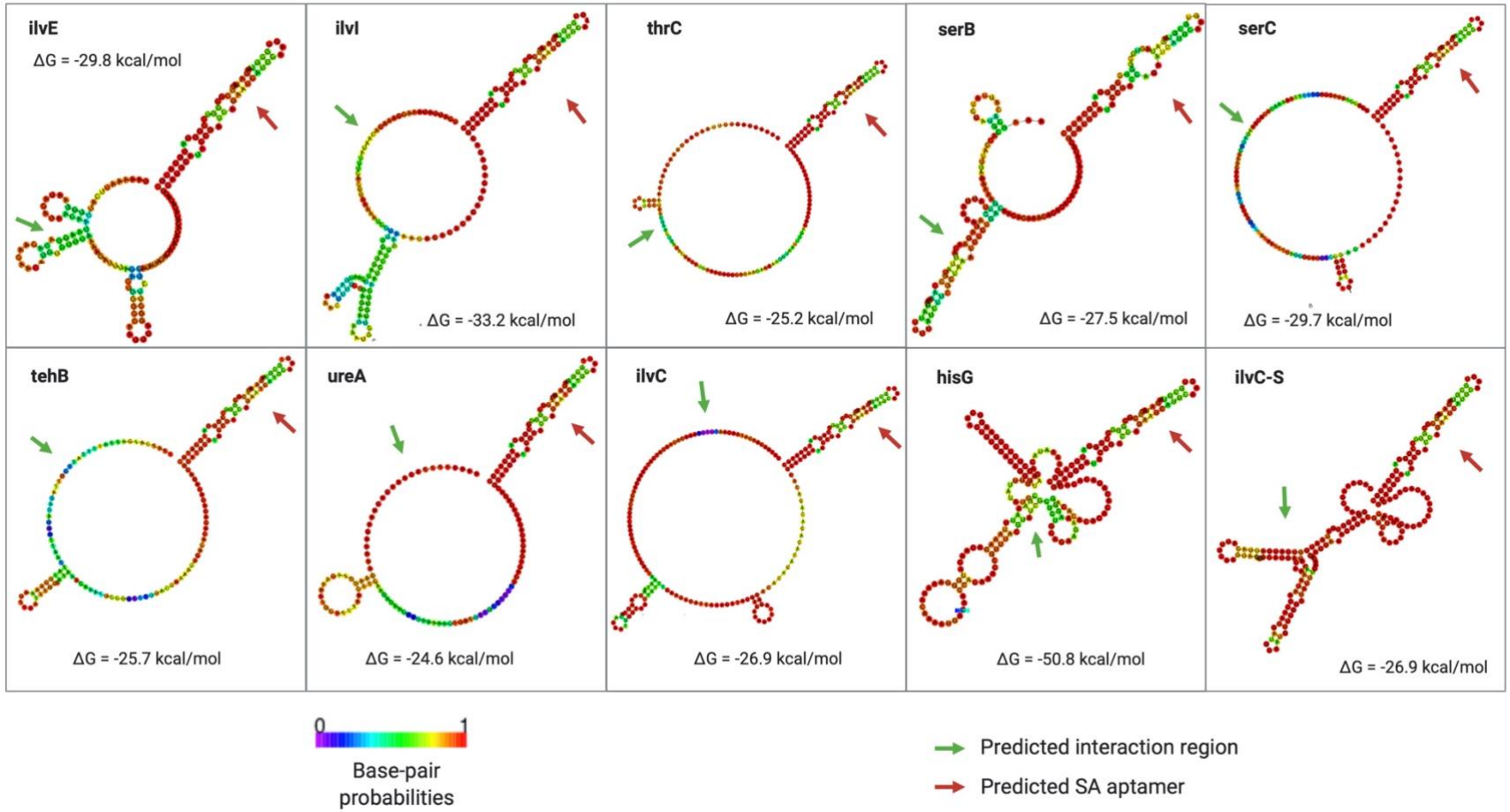
7. mRNA 5'UTR *ureA*	DNA template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGAAAATTTTACAAACACAAACAAAAGAAGGTACAAAATGCATTTAACTTCAAGAGAACAAGAAAAACTGATGTTGTTTCTTACACACACACACACACACACACGCATGCATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCCGGGATGCATGC *ureA* construct length = 193 MW(B): 119924 MW(F): 121121 GC = 41%	
	Primers:	**ureA FW1** Length = 81	TAATACGACTCACTATAGGGAAAATTTTACAAACACAAACAAAAGAAGGTACAAAATGCATTTAACTTCAAGAGAACAAG
		ureA FW2 Length = 80	GCATTTAACTTCAAGAGAACAAGAAAAACTGATGTTGTTTCTTACACACACACACACACACACACGCATGCATACCGACCAG
		SA Aptamer	GCATGCATCCCGGCCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGTATGCATGCGTG
8. mRNA 5'UTR *ilvC*	DNA template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGCATAACAACATAACAATGTGTAAACACAACATCATAAAACAACACACTCAACATCATACAATAAATAAGGAAGCACAAAATGGCTAACTATTTCAACACATTAAACTTACGTCAAAAATTAGACACACACACACACACACACACGCATGCATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCCGGGATGCATGC *ilvC* construct length = 236 MW(B): 146487 MW(F): 147684 GC = 40%	
	Primers:	**ilvC FW1** Length = 90	TAATACGACTCACTATAGGGCATAACAACATAACAATGTGTAAACACAACATCATAAAACAACACACTCAACATCATACAATAAATAAGG
		ilvC FW2 Length = 66	ACTCAACATCATACAATAAATAAGGAAGCACAAAATGGCTAACTATTTCAACACATTAAACTTACG
		ilvC RV1 Length = 74	CTGGTCGGTATGCATGCGTGTGTGTGTGTGTGTGTGTGTCTAATTTTTGACGTAAGTTTAATGTGTTGAAATAG
		SA Aptamer	GCATGCATCCCGGCCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGTATGCATGCGTG

9. mRNA 5'UTR *hisG*	DNA template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGTACACAACCTCTCGGAAGGCAACTTTCGAGAGGTTTTTTATATCCGAAATTTAAACACAACATGGGAAAAACAAAATGACAACCACAAACCGTTTACGTATCGCTCTGCAGAAGAAAGGGACACACACACACACACACACGCATGCATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCCGGGATGCATGC *hisG* construct length = 232 MW(B): 147118 MW(F): 148314 GC = 46%	
	Primers:	**hisG FW1** Length=90	TAATACGACTCACTATAGGGTACACAACCTCTCGGAAGGCAACTTTCGAGAGGTTTTTTATATCCGAAATTTAAACACAACATGG
		hisG FW2 Length=68	CCGAAATTTAAACACAACATGGGAAAAACAAAATGACAACCACAAACCGTTTACGTATCGCTCTGCAG
		hisG RV1 Length=66	CTGGTCGGTATGCATGCGTGTGTGTGTGTGTGTGTGTCCCTTTCTTCTGCAGAGCGATACGTAAACG
		SA Aptamer	GCATGCATCCCGGCCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGTATGCATGCGTG
10. mRNA 5'UTR *ilvC-S*	DNA template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGCATAACAACATAACAATGTATGATGGTGATGAGTAGATAACAACACACTCAACATCATACAATAAATAAGGAAGCACAAAATGGCTAACTATTTCAACACATTAAACTTACGTCAAAAATTAGACACACACACACACACACACGCATGCATACCGACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCCGGGATGCATGC *ilvC-S* construct length = 236 (scrambled areas in **bold**) MW(B): 147107 MW(F): 148303 GC = 41%	
	Primers:	**ilvC FW1** Length = 90	TAATACGACTCACTATAGGGCATAACAACATAACAATGTATGATGGTGATGAGTAGATAACAACACACTCAACATCATACAATAAATAAGG
		ilvC FW2 Length = 66	ACTCAACATCATACAATAAATAAGGAAGCACAAAATGGCTAACTATTTCAACACATTAAACTTACG
		ilvC RV1 Length = 74	CTGGTCGGTATGCATGCGTGTGTGTGTGTGTGTGTGTCTAATTTTTGACGTAAGTTTAATGTGTTGAAATAG
		SA Aptamer	GCATGCATCCCGGCCCGCGACTATCTTACGCACTTGCATGATTCTGGTCGGTATGCATGCGTG

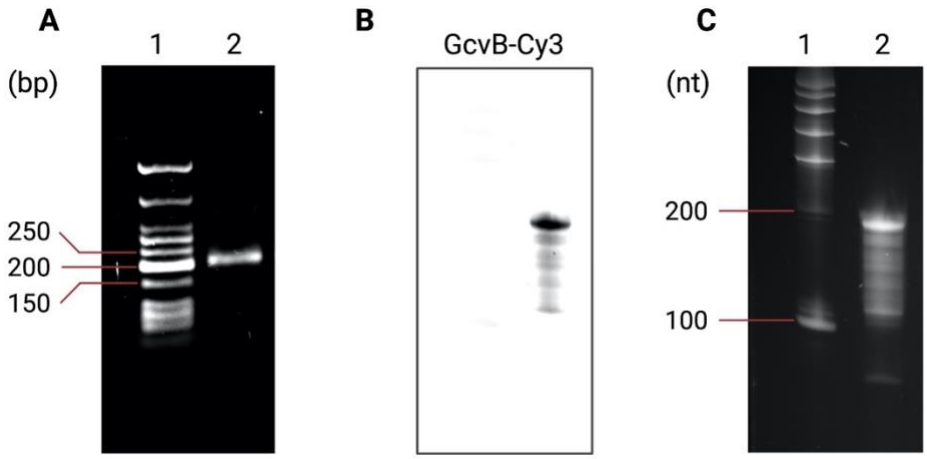
Designed DNA templates and associated primer sequences for preparing RNAs			
RNA name	**Sequences**		
1. sRNA *GcvB*	DNA template:	TAATACGACTCACTATAGGGTACTTAGACGGATAGTGATAGC TTAAGTTTCCGAATTATTACGGTAAATTCAGCTATTTCAGATTT TTAAGTATGATGTTGTGTTTGCATATTGGTCTAGGAAACTAGA CTGGAGTAACATCAAGTTACTCGTTTCACTTCCTGTATATTTTA AACCTTTTTGGTTTATAACCGTCCATTTTGGACGGTTTTTTT DNA MW: 132689 Length = 215 RNA MW: 63427 Length = 198	
	Primers:	**GcvB Fw1** Length=74	TAATACGACTCACTATAGGGTACTTAGACGGAT AGTGATAGCTTAAGTTTCCGAATTATTACGGT AAATTCAGC
		GcvB FW2 Length=70	CGAATTATTACGGTAAATTCAGCTATTTCAGAT TTTTAAGTATGATGTTGTGTTTGCATATTGGTC TAGGAAACTAGACTGGAG
		GcvB Rv1 Length=70	CCAAAAAGGTTTAAAATATACAGGAAGTGAAA CGAGTAACTTGATGTTACTCCAGTCTAGTTTCC TAGAC
		GcvB Rv2 Length=55	AAAAAAACCGTCCAAAATGGACGGTTATAAA CCAAAAAGGTTTAAAATATACAGG

Designed DNA templates and associated primer sequences for preparing RNAs			
RNA name	**Sequences**		
1. sRNA *GcvB-MG*	DNA template:	Biotin- **cgaggcagatct**TAATACGACTCACTATAGGGTACTTAGACGGATAGTGATAGCTTAAGTTTCCGAATTATTACGGTAAATTCAGCTATTTCAGATTTTTAAGTATGATGTTGTGTTTGCATATTGGTCTAGGAAACTAGACTGGAGTAACATCAAGTTACTCGTTTCACTTCCTGTATATTTTAAACCTTTTTGGTTTATAACCGTCCATTTTGGACGGTTTTTTTTCTCTCTCTCTCTCTCTCTCGGATCCCGACTGGCGAGAGCCAGGTAACGAATGGATCC DNA MW(B): 176757 Length = 285 RNA MW: 81903 Length = 256	
	Primers:	**GcvB-MG Fw1** Length=74	TAATACGACTCACTATAGGGTACTTAGACGGATAGTGATAGCTTAAGTTTCCGAATTATTACGGTAAATTCAGC
		GcvB-MG Fw2 Length=84	CGAATTATTACGGTAAATTCAGCTATTTCAGATTTTTAAGTATGATGTTGTGTTTGCATATTGGTCTAGGAAACTAGACTGGAG
		GcvB-MG Rv1 Length=70	CCAAAAAGGTTTAAAATATACAGGAAGTGAAACGAGTAACTTGATGTTACTCCAGTCTAGTTTCCTAGAC
		GcvB-MG Rv2 Length=59	GAGAAAAAAAACCGTCCAAAATGGACGGTTATAAACCAAAAAGGTTTAAAATATACAGG
		GcvB-MG Rv3 Length=77	GGATCCATTCGTTACCTGGCTCTCGCCAGTCGGGATCCGAGAGAGAGAGAGAGAGAGAAAAAAAAACCGTCCAAAATG

Appendix 10 – Secondary structure prediction for mRNA targets with attached SA aptamer.



Appendix 11 – PCR-TBIO and IVT of GcvB, fluorescently labelled by IVT with Cyanine 3 (Cy3) UTP.

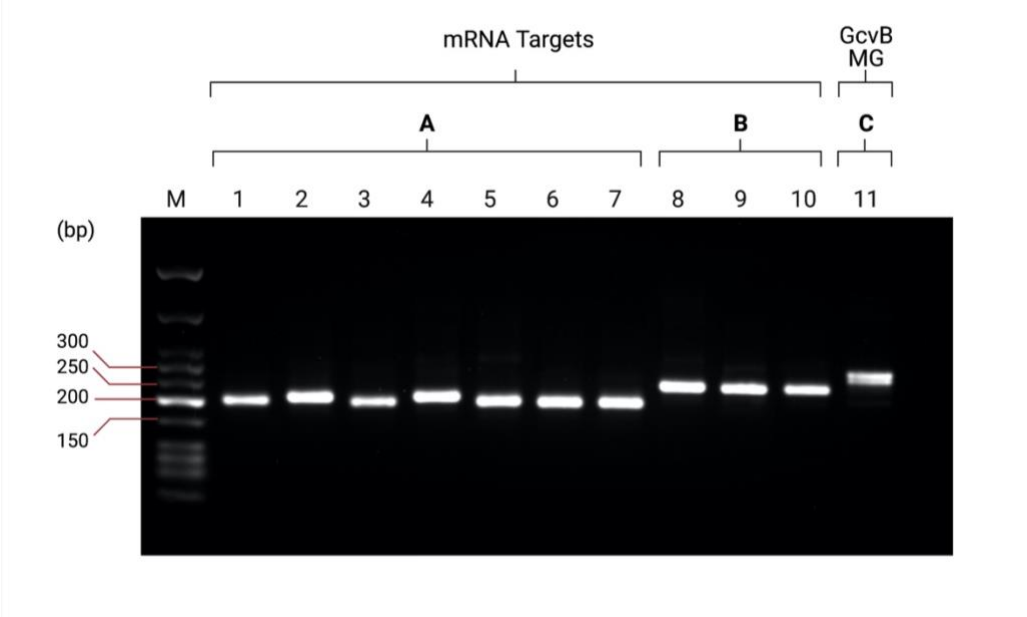


A) A 1.8% (w/v) agarose gel run in 1 x TBE containing lane (1): Low Molecular Weight DNA Ladder, lane (2) DNA IVT template for GcvB, visualised under a Syngene UV transilluminator

B) A pre-staining 8% (w/v) denaturing urea-PAGE gel run in 1 x TBE containing product of the IVT of the GcvB DNA IVT template from A, fluorescently labelled with Cy3 UTP, visualised by a Fuji FLA5000 phosphorimager using the Cy3 channel.

C) An 8% (w/v) denaturing Urea-PAGE gel stained with SYBR Gold™ containing lane (1): Low Range RNA Ladder, lane (2): product of the IVT of the GcvB DNA IVT template from A, fluorescently labelled with Cy3 UTP, visualised by a Fuji FLA5000 phosphorimager using the Cy3 channel.

Appendix 12 – PCR-TBIO of the DNA IVT templates for the predicted mRNA targets and the sRNA **M**odified **G**cvB (GcvB-**MG**).



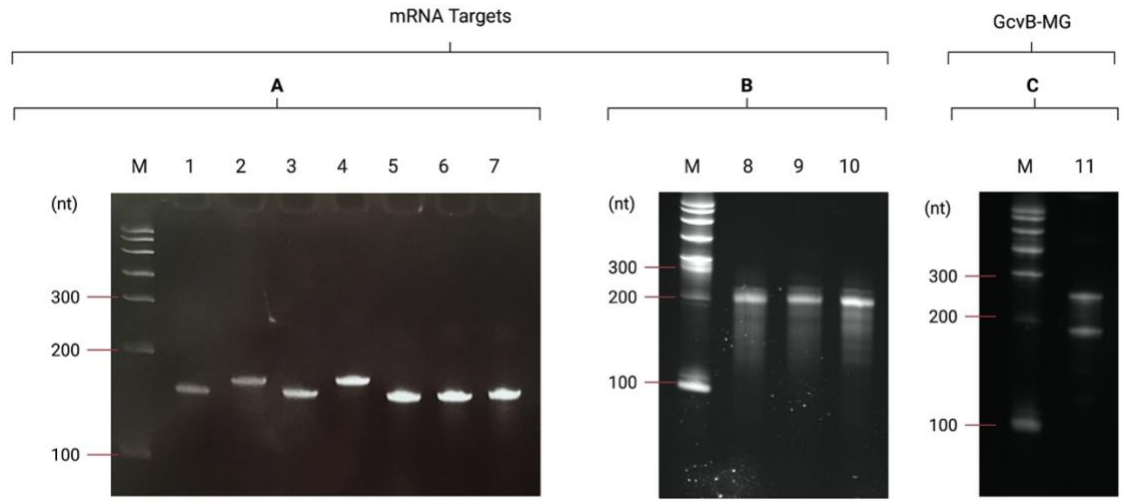
A 1.8% (w/v) agarose gel run in 1 x TBE, visualised under a Syngene UV transilluminator, containing lane (M) Low Molecular Weight DNA Ladder, and:

A) Construct A DNA IVT templates with 5 overlapping oligonucleotides: ilvE (1), ilvI (2), thrC (3), serB (4), serC (5), tehB (6), ureA (7).

B) Construct B DNA IVT templates with 6 overlapping oligonucleotides: ilvC (8), ilvC-S (9) and hisG (10).

C) DNA IVT templates with 6 overlapping oligonucleotides for GcvB-MG (11).

Appendix 13 – IVT product of DNA templates for the predicted mRNA targets and the sRNA **M**odified **G**cvB (GcvB-**MG**).



An 8% (w/v) denaturing Urea-PAGE gel run in 1 x TBE, gel stained with SYBR Gold™, and visualised under a Syngene UV transilluminator, containing lane (M): Low Range RNA Ladder, and:

A) IVT products of the respective DNA IVT templates in Appendix 9 (A): ilvE (1), ilvI (2), thrC (3), serB (4), serC (5), tehB (6), ureA (7).

B) IVT products of the respective DNA IVT templates in Appendix 9 (B): ilvC (8), ilvC-S (9) and hisG (10).

C) IVT product of the respective DNA IVT template for GcvB-MG (11). Note that this template two bands, one just under the 200 nucleotide (nt) mark, possibily related to the non-modified GcvB size (196 nt), and the other just under the 300 nt with the correct size for Gcvb-MG (288 nt). This may occur due to the RNA Polymerase detaching when reaching the termination signal in the original GcvB structure.

Appendix 14 – Buffer recipes.

10 x TBE (pH 8.4)	
Component	**Concentration**
Tris	0.89 M
Borate	0.89 M
EDTA	0.02 M

10 x PBS (pH 7.4)	
Component	**Concentration**
NaCl	1.37 M
KCl	0.027 M
Phosphate buffer	0.119 M

1 x PBST	
Component	**Concentration**
PBS	1 x
Tween-20 (pH 7.4)	0.02% (v/v)

Amine blocking buffer	
Component	**Concentration**
Ethanolamine (Fisher M251-1)	50 mM
PBST (pH 8.5 with HCl)	1 x

20 x SSC (pH 7.0)	
Component	**Concentration**
NaCl	3 M
Trisodium citrate	300 mM

Hybridisation buffer	
Component	**Concentration**
Tris (pH 7.8)	40 mM
$MgCl_2$	6 mM
NaCl	20 mM

Appendix 15 – Methods in Molecular Biology - *Manuscript in press*.

Generation of functional-RNA arrays by *in vitro* transcription and *in situ* RNA capture for the detection of RNA-RNA interactions

(Running head: Generation of functional-RNA arrays)

Helen A. Vincent, Charlotte A. Henderson, Daniela Lopes Cardoso and Anastasia J. Callaghan*

School of Biological Sciences and Institute of Biological and Biomedical Sciences, University of Portsmouth, Portsmouth, PO1 2DY, United Kingdom

* Corresponding author. Anastasia.Callaghan@port.ac.uk

Abstract

RNA performs a wide variety of vital cellular functions. These functions typically require interactions with other biological macromolecules, often as part of an intricate communication network. High-throughput techniques capable of analysing RNA-based interactions are therefore essential. Functional-RNA arrays address this need, providing the capability of performing hundreds of miniature assays in parallel. Here we describe a method to generate functional-RNA arrays using *in vitro* transcription of a DNA template array and *in situ* RNA capture. We also suggest how functional-RNA arrays could be applied to investigating RNA-RNA interactions.

Key words:

Binding partner; Functional-RNA; *In situ* capture; *In vitro* transcription; Microarray; RNA array; RNA-RNA interactions; Surface immobilisation;

1. Introduction

RNA is a structurally and functionally diverse molecule. Our understanding of its function(s) has transitioned from it playing relatively passive roles in translation (mRNA, rRNA and tRNA) to it performing a myriad of catalytic and/or regulatory activities (e.g. ribozymes, snRNAs, snoRNAs, miRNAs, riboswitches, sRNAs, lncRNAs, CRISPR) (Cech & Steitz, 2014). In many cases, it is known that regulatory activity of an RNA depends on interactions between the RNA and other molecules, for example complementary base-pairing with nucleic acids (RNA and DNA), formation of ribonucleoprotein complexes, or binding of small molecules (Marz & Stadler, 2011). These interactions, either in isolation or as part of a wider network, typically affect transcription, translation and/or RNA processing. Understanding the mechanistic details of the regulatory activities of RNA is critical for understanding both normal cellular function and disease, and has implications for therapeutic and synthetic biology applications. This in turn requires high-throughput techniques that are capable of detecting RNA-based interactions and analysing their functional outputs at the molecular level.

Microarray technology (Li *et al*., 2016), which can allow thousands of miniature assays to be performed in parallel on a single surface, has immense potential in this regard. RNA microarrays are in their infancy compared to DNA microarrays but, recently, a number of methods have been developed for generating RNA microarrays (Lietard and Somoza, 2019). These can be broadly classified into spotting, "on-array" transcription and *in situ* chemical synthesis (Lietard and Somoza, 2019). Applications such as detecting RNA-RNA (Phillips *et al*., 2018), small molecule-RNA (Phillips *et al*., 2018; Henderson *et al*., 2019) and protein-RNA (Beunrostro *et al*., 2014; She *et al*., 2017; Tome *et al*., 2014) interactions; monitoring regulatory outputs e.g. protein expression (Norouzi *et al*., 2019); and assaying RNA cleavage (Wu *et* al., 2014) are beginning to be explored.

"On-array" transcription methods for producing RNA arrays (Lietard and Somoza, 2019) are especially useful when arrays of longer, structured, functional RNAs are required rather than arrays of

shorter oligonucleotides. In this chapter, we provide the detailed protocols that are necessary for generating functional-RNA microarrays through an "on-array" method that involves *in vitro* transcription of a DNA *in vitro* transcription template array and *in situ* surface capture of the RNA on a facing surface (Phillips *et al*., 2018) (Fig. 1). As shown schematically in Fig. 1, this method first involves producing a DNA *in vitro* transcription template array by spotting custom-designed DNA *in vitro* transcription templates onto a microarray slide. A DNA *in vitro* transcription template array – *in vitro* transcription reagent mix – RNA capture surface "sandwich" is then assembled. As *in vitro* transcription proceeds, RNA synthesised from each DNA *in vitro* transcription template is captured *in situ* on an RNA capture surface to generate a corresponding functional-RNA array. The specific protocols that we include here are: *in vitro* transcription template design, generation of the DNA *in vitro* transcription template array by spotting, and generation of the functional-RNA array by *in vitro* transcription and *in situ* capture of the RNA.

[Fig. 1 near here]

We have utilised functional-RNA arrays that were generated following these protocols to detect regulatory RNA-RNA interactions (Phillips *et al*., 2018; Norouzi *et al*., 2019) and have investigated both RNA-RNA binding (Phillips *et al*., 2018) and the functional consequence of RNA-RNA interactions on protein expression (Norouzi *et al*., 2019). Since these applications have been reported elsewhere, and because the application of a functional-RNA array will be highly user-specific, we will not discuss the applications of RNA arrays in detail here. However, as an example, we will include a basic protocol for probing a functional-RNA array with a single binding partner in order to evaluate the specificity of an RNA-RNA binding interaction. This could be useful if a particular RNA has many RNA binding targets. For example, the regulatory RNA-RNA binding interactions between mRNAs and sRNAs, where a functional-RNA array of mRNA binding targets could be probed with an sRNA. Alternatively, it could be useful to investigate the importance of key nucleotides within a binding site, where a functional-RNA array of RNA binding target mutants could be probed with the binding partner.

2. Materials

To minimise the risk of RNA degradation, gloves should be worn when handling reagents, materials and equipment, and all solutions should be prepared using nuclease-free water (ultrapure water (18.2 MΩ/cm) filtered through a 0.22 μm filter).

2.1. *In vitro* transcription template design and synthesis

1. RNA secondary structure prediction software (*see* **Note 1**).

2.2 Generation of the DNA *in vitro* transcription template array

1. Streptavidin-coated glass microarray slide (*see* **Notes 2-4**).

2. 5′-biotinylated DNA *in vitro* transcription templates in phosphate-buffered saline (PBS), pH 7.4 (*see* **Notes 5-7**).

3. Automated arrayer (*see* **Note 8**) or a micropipette.

4. PBST: PBS, pH 7.4 supplemented with 0.05% (v/v) Tween 20.

5. PBS.

6. H_2O.

7. Microarray slide scanner (*see* **Notes 9** and **10**).

2.3 Generation of the functional-RNA array

1. DNA *in vitro* transcription template array (*see* **Methods 3.2**).

2. Streptavidin-coated glass microarray slide (*see* **Notes 2-4**).

3. Parafilm.

4. T7 *in vitro* transcription reagent mix (*see* **Note 11**): 1X MEGAscript T7 Transcription Reaction Buffer, 1X MEGAscript T7 Transcription Enzyme Mix, 0.4 mM ATP, 0.4 mM CTP, 0.4 mM GTP, 0.4 mM UTP.

5. PBST.

6. PBS.

7. H_2O.

8. Linker probe in 2X saline-sodium citrate (SSC) supplemented with 0.1% w/v sodium dodecyl sulphate (SDS) (*see* **Notes 12-14**).

9. LifterSlips (Thermo Fisher Scientific) (*see* **Note 15**).

10. Microarray slide scanner (*see* **Notes 9** and **10**).

2.4 Application of functional-RNA arrays to the evaluation of RNA-RNA binding specificity

1. Functional-RNA array (*see* **Methods 3.3**).

2. Fluorescently labelled binding partner probe in hybridisation buffer (*see* **Notes 16-18**).

3. LifterSlips (Thermo Fisher Scientific) (*see* **Note 15**).

4. Microarray slide scanner (*see* **Notes 9** and **10**).

3. Methods

The protocols provided below assume that the user has experience working in a molecular biology laboratory and is familiar with standard molecular biology techniques. The protocols are intended to provide a starting point when generating functional-RNA arrays for the first time. It is anticipated that the user will bring their system-specific knowledge to these protocols and will adapt them where appropriate.

Methods 3.1 involves the custom design and synthesis of *in vitro* transcription templates. This will require forward planning. Once complete, **Methods 3.2-3.4** can be completed in one day or they can be split over two days with **Methods 3.2** completed on the first day and **Methods 3.3** and **3.4** completed on the second day.

3.1 *In vitro* transcription template design and synthesis (*see* **Note 19**)

Each custom DNA *in vitro* transcription template should be designed to have the general architecture depicted in **Fig. 2**.

[Fig. 2 near here]

1. Generate the DNA *in vitro* transcription template sequence by linking the DNA sequence for each of the following components in order from 5′ to 3′: immobilisation linker (*see* **Note 20**) – promoter (*see* **Note 21**) – transcription system-specific sequence (*see* **Note 22**) – RNA of interest (user-defined) – unstructured linker (*see* **Note 23**) – immobilisation aptamer (*see* **Notes 24** and **25**) – fluorophore (*see* **Note 26**). A suggested starting DNA sequence for each component, except for the user-defined RNA of interest, is given in **Fig. 2**.

2. Generate RNA sequences for (i) the full-length transcribed RNA (transcription system-specific sequence – RNA of interest – unstructured linker – immobilisation aptamer) (*see* **Note 27**), (ii) the RNA of interest and (ii) the immobilisation aptamer.

[Fig. 3 near here]

3. Use RNA secondary structure prediction software (*see* **Note 1**) to predict the secondary structure for each of the RNA sequences generated in **Step 2** (*see* **Fig. 3** for an example).

4. Compare the predicted secondary structures for the RNA of interest and the immobilisation aptamer within the full-length transcript to their predicted secondary structures when they are in isolation to ensure that they are equivalent (*see* **Fig. 3**) (*see* **Note 28**).

5. Synthesise/source the DNA *in vitro* transcription templates (*see* **Note 29**).

3.2 Generation of the DNA *in vitro* transcription template array

1. Thaw a streptavidin-coated glass microarray slide at room temperature for 30 minutes (*see* **Notes 4** and **30**).

2. Spot the DNA *in vitro* transcription templates onto a streptavidin-coated microarray slide in the desired array layout (*see* **Notes 31** and **32**) (*see* **Fig. 4**).

[Fig. 4 near here]

2. Incubate in a humidified environment for 30 minutes at room temperature (*see* **Notes 33** and **34**).

3. Dip the DNA *in vitro* transcription template array into a 50 ml Falcon tube containing PBST (*see* **Note 35**).

4. Transfer the DNA *in vitro* transcription template array to a 50 ml Falcon tube containing 45 ml PBST (*see* **Note 35**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

5. Transfer the DNA *in vitro* transcription template array to a 50 ml Falcon tube containing 45 ml PBS (*see* **Note 35**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

6. Transfer the DNA *in vitro* transcription template array to a 50 ml Falcon tube containing 45 ml H_2O (*see* **Note 35**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

7. Dry the microarray slide by centrifugation at 500 x g for 5 min at room temperature (*see* **Note 36**).

8. If the DNA *in vitro* transcription templates are fluorescently labelled, visualise the array using a microarray slide scanner (*see* **Notes 26** and **37**).

9. Use the DNA *in vitro* transcription template array immediately for **Methods 3.3** or store at -20°C (*see* **Note 38**).

3.3 Generation of the functional-RNA array

1. Thaw a streptavidin-coated glass microarray slide (*see* **Notes 4** and **30**) and the DNA *in vitro* transcription template array generated in **Methods 3.2** (*see* **Notes 35** and **38**) at room temperature for 30 minutes.

2. Assemble a DNA *in vitro* transcription template array – *in vitro* transcription reagent mix – RNA capture surface "sandwich" as shown in **Fig. 5**. First place a streptavidin-coated RNA capture microarray slide surface-side up (*see* **Notes 30** and **39**). Place small pieces of parafilm on the short edges of the streptavidin-coated RNA capture microarray slide (*see* **Note 40**) (*see* **Fig. 5**). and pipette 90 μl *in vitro* transcription reagent mix on the RNA capture surface (*see* **Note 41**). Carefully place the DNA *in vitro* transcription template array (*see* **Note 35**) on top of the RNA capture slide array surface-side down.

[Fig. 5 near here]

3. Incubate at 37°C for 1 hr in a humidified environment (*see* **Note 42**).

4. Separate the DNA *in vitro* transcription template array slide and the functional-RNA array slide (*see* **Notes 43** and **44**).

5. Discard the DNA *in vitro* transcription template array (*see* **Note 45**).

6. Transfer the functional-RNA array slide to a 50 ml Falcon tube containing 45 ml PBST (*see* **Note 44**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

7. Transfer the functional-RNA array slide to a 50 ml Falcon tube containing 45 ml PBS (*see* **Note 44**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

8. Transfer the functional-RNA array slide to a 50 ml Falcon tube containing 45 ml H_2O (*see* **Note 44**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

9. Dry the functional-RNA array slide by centrifugation at 500 x g for 5 min at room temperature (*see* **Note 36**).

10. Probe the functional-RNA array immediately (*see* below).

[Fig. 6 near here]

11. Place the functional-RNA array with the array surface-side up (*see* **Note 44**) and pipette 90 μl linker probe (*see* **Notes 12-14** and **46**) onto the functional-RNA array and cover with a 24 x 60 mm LifterSlip (*see* **Fig. 6**).

12. Incubate at room temperature for 30 minutes in the dark (*see* **Note 47**).

13. Remove the LifterSlip, transfer the functional-RNA array slide to a 50 ml Falcon tube containing 45 ml PBST (*see* **Note 44**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

14. Transfer the functional-RNA array slide to a 50 ml Falcon tube containing 45 ml PBS (*see* **Note 44**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

15. Transfer the functional-RNA array slide to a 50 ml Falcon tube containing 45 ml H_2O (*see* **Note 44**) and rotate the tube on a rolling platform for 5 minutes at room temperature.

16. Dry the functional-RNA array slide by centrifugation at 500 x g for 5 min at room temperature (*see* **Note 36**).

17. Visualise the functional-RNA array using a microarray slide scanner (*see* **Note 48**).

3.4 Application of functional-RNA arrays to the evaluation of RNA-RNA binding specificity

[Fig. 7 near here]

1. Design the functional-RNA array layout (*see* **Notes 32** and **49**) (*see* **Fig. 7**).

2. Design each *in vitro* transcription template that will be needed to generate the functional-RNA array (*see* **Methods 3.1**).

3. Generate the DNA *in vitro* transcription template array (*see* **Methods 3.2**).

4. Generate the functional-RNA array (*see* **Methods 3.3**). Use immediately.

5. Probe the functional-RNA array with the RNA binding partner probe (*see* **Notes 16-18** and **51**). (i) Place the functional-RNA array with the array surface-side up (*see* **Note 44**) and pipette 90 µl RNA binding partner probe (*see* **Notes 16-18** and **52**) onto the functional-RNA array and cover with a 24 x 60 mm LifterSlip (*see* **Note 53**). (ii) Incubate for 30 minutes at room temperature in the dark (*see* **Note 54**). (iii) Remove the LifterSlip, transfer the functional-RNA array slide to a 50 ml Falcon tube containing 45 ml hybridisation buffer (*see* **Note 44**) and rotate on a rolling platform for 5 minutes at room temperature. (iv) Repeat step (iii) two more times (*see* **Note 55**). (v) Dry the functional-RNA array slide by centrifugation at 500 x g for 5 min at room temperature (*see* **Note 36**).

6. Visualise the functional-RNA array using a microarray slide scanner (*see* **Note 56**).

7. Quantify and analyse the data (*see* **Fig. 7**).

4. Notes

1. We use RNAfold from the ViennaRNA Package 2.0 (rna.tbi.univie.ac.at) (Lorenz *et al*., 2011) which is freely available online as a web interface.

2. Streptavidin-coated glass microarray slides are commercially available e.g. Nexterion HS slides (Schott). However, we covalently immobilise streptavidin to a Nexterion H slide (Schott) using amine coupling to generate our streptavidin-coated surface. This allows us to control the density of streptavidin on the surface and we achieve better results when we have a lower density of streptavidin on the DNA capture surface than on the RNA capture surface. A protocol for generating streptavidin-coated surfaces in this manner is provided in **Note 3**.

3. (i) Pipette 90 µl of 1 µM streptavidin (Sigma) in PBS, pH 7.4 (DNA capture surface) or 16.7 µM streptavidin in PBS, pH 7.4 (RNA capture surface) onto a Nexterion H slide (Schott). Cover with a 24 x 60 mm LifterSlip (Thermo Fisher Scientific) and incubate at 37°C for 1 hour in a humidified environment (we use a mini incubator (9.2 L) and a Petri dish containing 25 ml H_2O). (ii) Remove the LifterSlip, place the microarray slide in a 50 ml Falcon tube containing 45 ml PBST and rotate the tube on a rolling platform for 5 minutes at room temperature. (iii) Transfer the microarray slide to a 50 ml Falcon tube containing 45 ml PBS and rotate the tube on a rolling platform for 5 minutes at room temperature. (iv) Transfer the microarray slide to a 50 ml Falcon tube containing 45 ml H_2O and rotate the tube on a rolling platform for 5 minutes at room temperature. (v) Transfer the microarray slide to a 50 ml Falcon tube containing 45 ml 50 mM ethanolamine-HCl, pH 8.5 (to block any unreacted NHS functional groups) and rotate the tube on a rolling platform for 30 minutes at room temperature. (vi) Repeat the washes in steps (ii)-(iv). (vii) Dry the microarray slide by centrifugation at 500 x g for 5 min at room temperature (we place the slide in a 50 ml Falcon tube and use a swing bucket centrifuge). (viii) Use the prepared capture surface immediately (*see* **Methods 3.2-3.3**) or store at -20°C for up to 12 months.

4. Streptavidin-coated glass microarray slides are typically stored at -20°C and should be thawed at room temperature for 30 minutes prior to use.

5. The DNA *in vitro* transcription templates must be custom designed and synthesised (*see* **Methods 3.1** and **Fig. 2**).

6. When working with a DNA *in vitro* transcription template for the first time, we typically prepare it, and test it, at a range of concentrations between 20 nM and 500 nM to determine the optimal concentration. If it is only possible to test a single concentration, 200 nM is a good starting point.

7. If the DNA *in vitro* transcription templates are labelled at the 3′ end with a fluorophore (*see* **Methods 3.1** and **Fig. 2**) they should be kept in the dark, either by using opaque tubes or wrapping the tubes in foil.

8. We use a Qarray2 (Genetix).

9. We use a GenePix 4300A microarray scanner with integrated GenePix Pro 7 image analysis software (Molecular Devices). The scanner has three lasers that allow excitation at 488 nm, 532 nm or 635 nm and three emission filters: Standard Blue (534/42 nm), Standard Green (579/34 nm) and Standard Red (676/29 nm).

10. Ensure that the fluorescence properties (excitation and emission wavelengths) of the chosen fluorophore(s) are compatible with the excitation and emission capabilities of the available microarray slide scanner.

11. We use the MEGAscript T7 Transcription Kit (Invitrogen) to assemble the T7 *in vitro* transcription reagent mix. However, alternative kits could be substituted or the reagent mix could be assembled from individually sourced components.

12. The linker probe is a fluorescently labelled DNA oligonucleotide with a sequence complementary to the unstructured linker between the RNA of interest and the immobilisation aptamer (*see* **Methods 3.1** and **Fig. 2**). For the 5′-ACA CAC ACA CAC ACA CAC AC-3′ linker we use a 5′-Dy649-GTG TGT GTG TGT GTG TGT GT-3′ DNA oligonucleotide linker probe. When designing the linker probe and selecting the fluorophore, care should be taken to ensure that its fluorescence properties are compatible with the excitation and emission capabilities of the available microarray slide scanner.

13. We typically prepare and use the linker probe at a concentration between 50 nM and 100 nM.

14. The fluorescently labelled linker probe should be kept in the dark, either in opaque tubes or in tubes wrapped in foil.

15. We routinely use 24 x 60 mm LifterSlips to cover the full microarray slide or 22 x22 mm LifterSlips to cover approximately 1/3 of the microarray slide.

16. The functional binding interaction to be detected will be highly user-specific. For RNA-RNA interactions the RNA binding partner probe should be a fluorescently labelled RNA containing the required elements for binding (e.g. a specific nucleotide sequence and/or a specific RNA structural element). The RNA binding probe should be labelled with a fluorophore with orthogonal fluorescence properties to that used for the linker probe. This will allow simultaneous/serial binding and detection of both the linker probe and the RNA binding partner. We use a standard method to produce internally fluorescently labelled RNA binding partner probes by *in vitro* transcription, supplementing the *in vitro* transcription mix with Cy3/5-UTP (GE Healthcare Life Sciences). When using the Dy649 linker probe (*see* **Note 12**) we would label our RNA binding partner probe with Cy3.

17. For the RNA-RNA interactions that we have experience with, we have found that 40 mM Tris, pH 7.8, 6 mM $MgCl_2$, 20 mM NaCl is a good hybridisation buffer. However, the hybridisation buffer composition should be adjusted to accommodate any system-specific requirements.

18. The required RNA binding partner probe concentration is likely to be system- and/or experiment-specific. For a yes/no output, we usually select an RNA binding partner probe concentration well above the expected K_d for the interaction in the first instance. If the K_d is unknown, a range of RNA binding partner probe concentrations spanning several orders of magnitude (e.g. 1 nM, 10 nM, 100 nM, 1 μM) should be tested.

19. This protocol should be followed for each custom DNA *in vitro* transcription template that is required.

20. We typically use a 5′-biotinylated linker to facilitate surface immobilisation of the *in vitro* transcription template when it is spotted onto a streptavidin-coated glass microarray slide. Alternative immobilisation chemistries, e.g. amine coupling, and/or surfaces could be employed with the appropriate modification(s) to these protocols (Li *et al.*, 2016).

21. We typically use the T7 promoter, which facilitates transcription by T7 RNA polymerase. The T7 *in vitro* transcription system is routinely used by molecular biologists and T7 *in vitro* transcription kits are commercially available at a reasonable cost. Alternative promoters and transcription systems could be employed with the appropriate modifications(s) to these protocols.

22. T7 RNA polymerase requires a G to be in the +1 position and a G in the +2 and +3 positions are required for optimal transcription efficiency (Imburgio *et al*., 2000). Depending on the sequence of the RNA of interest, we insert between 0 and 3 Gs between the T7 promoter and the RNA of interest to ensure that there is a G in positions +1, +2 and +3. On rare occasions, the addition of these nucleotides can adversely affect the structure of the RNA of interest (*see* **Note 28**). In this case, it is possible to omit the G from positions +2 and +3.

23. This is used to physically separate the RNA of interest and the immobilisation aptamer in order to promote the independent folding of each of these RNA modules. The linker can also be utilised as a binding site for a fluorescently labelled complementary DNA oligonucleotide linker probe. This enables visualisation of the RNA array and quantification of the relative RNA level at each position of the array (Norouzi *et al*., 2019; Henderson *et al*., 2019).

24. We typically use the streptavidin-binding RNA aptamer (SA_{apt}) to facilitate *in situ* capture of the *in vitro* transcribed RNA on a streptavidin-coated glass microarray slide. Alternative immobilisation systems may be employed with the appropriate modification(s) to these protocols e.g. the tobramycin-binding RNA aptamer and a tobramycin-coated surface (Phillips *et al*., 2018).

25. The immobilisation aptamer should always be incorporated 3′ to the RNA of interest to ensure that only fully-transcribed, full-length RNAs are captured.

26. Inclusion of a 3′ fluorophore is optional. However, we typically include a Dy549 fluorophore at the 3′ end of our DNA *in vitro* transcription templates to allow visualisation and quantification of the relative levels of DNA at each position on the *in vitro* transcription template array. This can be a useful quality control checkpoint when generating the DNA *in vitro* transcription template array and it can be used subsequently to calculate RNA transcription/capture efficiencies (Norouzi *et al*., 2019; Henderson *et al*., 2019). Alternative fluorophores to Dy549 may be employed (*see* **Note 10**).

27. RNA polymerase will begin transcribing from the system-specific sequence (GGG for T7 RNA polymerase) and continue to transcribe the RNA of interest, the unstructured linker and the immobilisation aptamer. The immobilisation linker and the promoter will not be transcribed.

28. Both the RNA of interest and the immobilisation aptamer must fold correctly to generate a functional-RNA array. Misfolding of the RNA of interest can result in a non-functional RNA that may be unsuitable for downstream applications and misfolding of the immobilisation aptamer can result in the failure of *in situ* RNA capture by the surface even if the RNA of interest is folded correctly. Therefore, it is critical to check that linking the RNA of interest to an immobilisation aptamer does not affect the folding of either RNA. Where secondary structure predictions indicate that correct folding of both RNAs within the context of the full-length transcript is likely, we accept the DNA *in vitro* transcription template design. However, where RNA secondary structure predictions indicate that misfolding of the immobilisation aptamer and/or the RNA of interest is likely, we typically repeat the design process with alternative linker and/or modified immobilisation aptamer sequences until RNA secondary structure predictions suggest that both are likely to fold correctly. The linker sequences and modified streptavidin aptamer sequences that we have successfully utilised are listed in **Table 1**.

[Table 1 near here]

29. DNA *in vitro* transcription templates can be synthesised using a variety of standard molecular biology methods e.g. gene synthesis, primer extension or annealing of complementary oligonucleotides and the choice of method will depend primarily on the length of the RNA of interest. We typically perform gene synthesis in-house. We use DNAWorks v3.2.4 (https://hpcwebapps.cit.nih.gov/dnaworks; Hoover and Lubkowski, 2002) to design gene synthesis primers and follow standard gene synthesis protocols (Gao *et al.*, 2003) to first produce a double-stranded non-biotinylated and non-fluorescently labelled template. We then amplify this template in a standard PCR reaction with a 5′-biotinylated forward primer and a 5′-Dy549 fluorescently labelled

reverse primer to produce the final 5′-biotinylated, 3′-fluorescently labelled DNA *in vitro* transcription template (Henderson *et al*., 2019). DNA *in vitro* transcription template preparation could also be outsourced to commercial gene synthesis services e.g. GeneArt (Thermo Fisher Scientific) or custom oligonucleotide manufacturers e.g. Integrated DNA Technologies.

30. Take care not to touch the streptavidin-coated surface when handling the microarray slides.

31. As *in vitro* transcription and *in situ* capture proceeds, there will be some level of diffusion prior to RNA capture which results in larger spots on the functional-RNA array compared to the DNA template array (Phillips *et al*., 2018). This is impacted by factors including *in vitro* transcription efficiency and incubation time, and spacing/volume between the slides. Spacing of the DNA spots on the DNA template array may need to be adjusted to account for this. We typically use an automated Qarray2 arrayer (Genetix) fitted with a 200 μm pinhead to spot the DNA *in vitro* transcription template onto a streptavidin-coated glass microarray slide at a spot separation of 1250 μm. This allows us to achieve a spot density of ~400 spots per slide. It is also possible to spot the DNA *in vitro* transcription template manually using a micropipette. If we manually pipette 0.2 μl per spot we can achieve a spot density of 24 spots per slide.

32. Note that the array layout of the functional-RNA array will be the mirror-image of the DNA *in vitro* template array layout.

33. If spotting is performed using an automated arrayer, the arrayer can provide the humidified environment. If spotting is performed manually with a micropipette, a humidified environment can be created using a mini incubator (9.2 L) and a Petri dish containing 25 ml H_2O.

34. If the DNA *in vitro* transcription templates are fluorescently labelled, they should be kept in the dark as much as possible during spotting. The Qarray2 automated arrayer has a tinted cover or, when manually spotting, we cover the window panel in the mini incubator with foil during the incubation step.

35. Take care not to touch the DNA *in vitro* transcription template array surface when handling the microarray slides.

36. We place the slide in a 50 ml Falcon tube and use a swing bucket centrifuge.

37. We visualise our array of 3′-Dy549 labelled DNA *in vitro* transcription templates using an excitation wavelength of 532 nm and a Standard Green (579/34 nm) emission filter with a GenePix 4300A microarray scanner (Molecular Devices) (*see* **Note 9**).

38. We typically store the DNA *in vitro* transcription template array at -20°C overnight although it can be stored at -20°C for at least seven days. The DNA *in vitro* transcription template array should be thawed at room temperature for 30 minutes prior to use.

39. It is important to assemble the DNA *in vitro* transcription template array – *in vitro* transcription reagent mix – RNA capture surface "sandwich" with the RNA capture surface as the lower layer so that the *in vitro* transcription reagent mix is pipetted onto this surface. This ensures that transcription is not initiated until the DNA *in vitro* transcription template array is added to complete the "sandwich assembly. This in turn helps to confine the *in vitro* transcription and *in situ* RNA capture to a discrete volume (*see* **Note 31**).

40. The use of parafilm as spacers helps with the separation of the DNA *in vitro* transcription template array and the functional-RNA array following *in vitro* transcription and *in situ* RNA capture. However, using spacers also increases the distance that the RNA must diffuse between synthesis and capture and can result in larger RNA spots (*see* **Note 31**). If a high spot density is required, it may be beneficial to omit the spacers and reduce the volume of *in vitro* transcription mix that is used.

41. We have used *in vitro* transcription mix volumes between 12 and 150 μl. 90 μl allows for good coverage of the array and for the use of spacers (*see* **Note 40**) to assist slide separation following *in vitro* transcription and *in situ* RNA capture.

42. We typically incubate at 37°C for between 30 and 90 minutes in a mini incubator (9.2 L) that has been humidified using a Petri dish containing 25 ml H_2O. For most applications, this is sufficient for *in vitro* transcription and *in situ* RNA capture. Longer incubation times may result in larger spots (*see* **Note 31**).

43. The increased surface tension of the *in vitro* transcription mix as the DNA *in vitro* transcription template array and the functional-RNA array are pulled apart can make it difficult to separate the

arrays, especially if parafilm spacers have not been used. It can be helpful to submerge the sandwich assembly in PBS when separating the arrays.

44. Take care not to touch the functional-RNA array surface when handling the microarray slides.

45. In theory, the DNA *in vitro* transcription template array slide could be washed and reused. We do not currently reuse the DNA *in vitro* transcription template array.

46. Probing the functional-RNA array with linker probe provides a quality control checkpoint when generating the functional-RNA array and it can be used to calculate RNA transcription/capture efficiencies (Norouzi *et al*., 2019; Henderson *et al*., 2019).

47. Since the linker probe is fluorescently labelled, the incubation should be performed in the dark. We cover the assembly with foil.

48. We visualise Dy649 labelled linker probe bound to our functional-RNA array using an excitation wavelength of 635 nm and a Standard Red (676/29 nm) emission filter with a GenePix 4300A microarray scanner (Molecular Devices) (*see* **Note 9**).

49. The functional-RNA array will be an array of functional-RNAs that will be probed with a single RNA binding partner probe (*see* **Notes 16** and **50**). The functional-RNA array should include a positive control RNA that the RNA binding partner probe is known to bind, a negative control RNA that the RNA binding partner probe is known not to bind, and test RNAs (*see* **Note 51**) that may or may not be bound by the RNA binding partner probe (*see* **Fig. 7**). We typically design a single field of between 9 and 16 spots. This field may contain a single RNA at different concentrations, with each field containing a different RNA, or multiple RNAs at the same concentration. It is good practice to include replicate spots of each RNA, and/or replicate fields.

50. It is possible to simultaneously probe with more than one RNA binding partner probe provided that each RNA binding partner probe can be labelled with an orthogonal fluorophore (Phillips *et al*., 2018).

51. The test RNAs may be potential binding targets of the RNA binding partner probe (e.g. in the case of an sRNA that has multiple mRNA targets) or they may be a series of binding-site mutants designed to evaluate the key residues required for a specific RNA-RNA interaction.

52. The probing conditions needed for the RNA binding partner probe may be highly system-specific. For example, it may be necessary to test different hybridisation buffers (*see* **Note 17**), different RNA binding partner probe concentrations (*see* **Note 18**), different incubation temperatures and/or times.

53. Different conditions and/or different RNA binding partner probes can be tested on the same array by using 20 μl RNA binding partner probe (*see* **Notes 16-18** and **51**) and 22 x 22 mm LifterSlips to cover a smaller area of the array. Up to three conditions can be tested per array in this manner.

54. Since the RNA binding parnter probe is fluorescently labelled, the incubation should be performed in the dark. We cover the assembly with foil.

55. Depending on the affinity of the RNA-RNA interaction and the non-specific binding observed, fewer or more wash steps may be desirable.

56. We visualise Cy3 labelled RNA binding partner probe bound to our functional-RNA array using an excitation wavelength of 532 nm and a Standard Green (578/34 nm) emission filter with a GenePix 4300A microarray scanner (Molecular Devices) (*see* **Note 9**).

5. References

Cech, TR and Steitz, JA (2014) The Noncoding RNA Revolution – Trashing Old Rules to Forge New Ones. *Cell*, **157**, 77-94

Marz, M and Stadler, PF (2011) RNA Interactions. In: Collins, LJ (ed) RNA Infrastructure and Networks. Advances in Experimental Medicine and Biology, vol. 722. Springer, New York

Li, PCH, Sedighi, A and Wang, L (eds) (2016) Microarray Technology. Methods in Molecular Biology, vol. 1368. Humana Press, New York

Lietard J and Somoza MM (2019) Spotting, Transcription and *In Situ* Synthesis: Three Routes for the Fabrication of RNA Microarrays. *Comput. Struct. Biotechnol. J.* **17**, 862-868

Phillips, JO, Butt, LE, Henderson, CA, Devonshire, M, Healy, J, Conway, SJ, Locker, N, Pickford, AR, Vincent, HA and Callaghan AJ (2018) High-density Functional-RNA Arrays as a Versatile Platform for Studying RNA-based Interactions. *Nucleic Acids Res.* **46**, e86

Henderson, CA, Rail, CA, Butt, LE, Vincent HA and Callaghan AJ (2019) Generation of Small Molecule-binding RNA arrays and Their Application to Fluorogen-binding RNA Aptamers. *Methods* **167**, 39-53

Buenrostro JD, Araya, CL, Chircus, LM, Layton, CJ, Chang, HY, Snyder, MP and Greenleaf, WJ (2014) Quantitative Analysis of RNA-protein Interactions on a Massively Parallel Array Reveals Biophysical and Evolutionary Landscapes. *Nat. Biotechnol.* **32**, 562-568

She, R, Chakravarty, AK, Layton, CJ, Chircus, LM, Andreasson, JO, Damaraju, N, McMahon, PL, Buenrostro JD, Jarosz, DF and Greenleaf, WJ (2017) Comprehensive and Quantitative Mapping of

RNA-protein Interactions Across a Transcribed Eukaryotic Genome. *Proc. Natl. Acad. Sci. USA* **114**, 3619-3624

Tome, JM, Ozer, A, Pagano, JM, Gheba, D., Schroth, GP and Lis JT (2014) Comprehensive Analysis of RNA-protein Interactions by High-throughput Sequencing-RNA Affinity Profiling. *Nat. Methods* **11**, 683-688

Norouzi, M, Pickford, AR, Butt, LE, Vincent, HA and Callaghan AJ (2019) Application of mRNA Arrays for the Production of mCherry Reporter-protein Arrays for Quantitative Gene Expression Analysis. *ACS Synth. Biol.* **8**, 207-215

Wu, H-C, Holden, MT and Smith LM (2014) Enzymatic Fabrication of High-density RNA Arrays. *Angew. Chem. Int. Ed.* **53**, 13514-13517

Imburgio, D, Rong, M, Ma, K and McAllister, WT (2000) Studies of Promoter Recognition and Start Site Selection by T7 RNA Polymerase Using a Comprehensive Collection of Promoter Variants. *Biochemistry* **39**, 10419-10430

Lorenz, R, Bernhart, SH, Höner zu Siederdissen, C, Tafer, H, Flamm, C, Stadler, PF and Hofacker, IL (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26

Hoover, DM and Lubkowski, J (2002) DNAWorks: An Automated Method for Designing Oligonucleotides for PCR-based Gene Synthesis. *Nucleic Acids Res.* **30**, e43

Gao, X, Yo, P, Keith, A, Ragan, TJ, Harris, TK (2003) Thermodynamically Balanced Inside-Out (TBIO) PCR-based Gene Synthesis: A Novel Method of Primer Design for High-fidelity Assembly of Longer Gene Sequences. *Nucleic Acids Res.* **31**, e143

Kerpedjiev, P, Hammer and S, Hofacker, IL (2015) Forna (Force-directed RNA): Simple and Effective Online RNA Secondary Structure Diagrams. *Bioinformatics* **31**, 3377-3379

Leppek, K, and Stoecklin, G (2014) An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* **42**, e13

Figure Captions

Fig. 1. Generation of functional-RNA arrays by *in vitro* transcription and *in situ* RNA capture. Custom DNA *in vitro* transcription templates are designed and synthesised. These are spotted onto a microarray slide to produce a DNA *in vitro* transcription template array. A DNA *in vitro* transcription template array – *in vitro* transcription reagent mix (IVT) – RNA capture surface "sandwich" is assembled. As *in vitro* transcription proceeds, RNA is captured *in situ* to produce a corresponding functional-RNA array.

Fig. 2. Design of the *in vitro* transcription template. A schematic of the *in vitro* transcription template which consists of a short immobilisation linker (*see* **Note 20**), a promoter (*see* **Note 21**), transcription system-specific sequence (*see* **Note 22**), sequence encoding the RNA of interest, an unstructured linker (*see* **Note 23**), sequence encoding an immobilisation aptamer (*see* **Notes 24** and **25**) and an optional fluorophore (*see* **Note 26**). A suggested starting DNA sequence for each component of the *in vitro* transcription template is indicated, except for the user-specified RNA of interest sequence which is represented by **…$(N)_n$….**

Fig. 3. Predicted secondary structures for an RNA of interest (MicA) transcript. (A) The sequence of the full-length RNA transcript which consists of GGG (T7 transcription system-

specific sequence; grey), MicA (a small regulatory RNA of interest; blue), an unstructured linker (magenta) and the streptavidin-binding RNA aptamer (SA_{apt}; immobilisation aptamer; gold). The RNA secondary structures, predicted by RNAfold (rna.tbi.univie.ac.at; Lorenz *et al*., 2011), for the full-length RNA transcript, MicA and the streptavidin-binding aptamer are shown below the RNA sequence in dot-bracket notation. (B)-(D) The RNA secondary structures, predicted by RNAfold, visualised in forna (Kerpedjiev *et al*., 2015). The predicted secondary structures for MicA and the streptavidin-binding aptamer are the same in the context of the full-length transcript and in isolation.

Fig. 4. Spotting and visualisation of the DNA *in vitro* transcription template array. The DNA *in vitro* transcription template array is prepared by spotting 5′-biotinylated DNA *in vitro* transcription templates onto a streptavidin-coated microarray slide. Spotting can be performed manually using a micropipette, or it can be automated using an arrayer (*see* **Note 30**). Both methods result in discrete spots containing multiple immobilised DNA *in vitro* transcription template molecules. Following spotting, the DNA *in vitro* transcription template array is washed. If the DNA *in vitro* transcription templates are also fluorescently labelled, the DNA *in vitro* transcription template array can be visualised using a microarray scanner and quantified (*see* **Note 37**).

Fig. 5. *In vitro* transcription and *in situ* RNA capture. (A) Assembly of the DNA *in vitro* transcription template array – *in vitro* transcription reagent mix (IVT) – RNA capture surface "sandwich". The RNA capture surface (a streptavidin-coated microarray slide) is position surface-side up. A small piece of parafilm is positioned at each of the short edges of the RNA capture surface to act as spacers. *In vitro* transcription reagent mix is pipetted onto the RNA capture surface and the DNA *in vitro* transcription template array, array-side down, is carefully lowered onto the RNA capture surface to complete the "sandwich" assembly. (B) The "sandwich" assembly is incubated at 37°C for 30-90 minutes during which time RNA is synthesised by *in vitro*

transcription and captured *in situ* by the RNA capture surface. (C) The DNA *in vitro* transcription template array and the newly generated corresponding functional-RNA array are carefully separated.

Fig. 6. Probing the functional-RNA array with linker probe. The functional-RNA array can be probed with a fluorescently labelled DNA oligonucleotide with a sequence complementary to the unstructured linker region between the RNA of interest and the streptavidin-binding RNA aptamer (*see* **Note 12**). This allows visualisation and quantification of the RNA levels on the functional-RNA array (*see* **Note 46**). A solution of fluorescently labelled linker probe is pipetted over the functional-RNA array (*see* **Note 13**). This is covered with a LifterSlip and incubated at room temperature for 30 minutes in the dark (*see* **Note 47**). During incubation, the linker probe binds to the unstructured linker on the immobilised functional-RNAs. Following incubation, the LifterSlip is removed, the functional-RNA array is washed and linker probe fluorescence is detected using a microarray scanner and quantified (*see* **Note 48**).

Fig. 7. Application of functional-RNA arrays to the evaluation of RNA-RNA binding specificity. Functional-RNA arrays can be used to detect RNA-RNA interactions (Phillips *et al*., 2018). (A) A functional-RNA array is designed with appropriate positive (+) and negative (-) control RNAs and a variety of test RNAs (T1-T7) (*see* **Note 49**). (B) A DNA *in vitro* transcription template array is generated (*see* **Note 16**). (C) A functional-RNA array is generated by *in vitro* transcription and *in situ* RNA capture and probed with linker probe (*see* **Notes 12-14**, **32** and **46**). (D) The functional-RNA array is probed with RNA binding partner probe (*see* **Notes 16-18**, **32** and **51**). Fluorescence should be clearly detected for the positive control and not for the negative control. The presence or absence of fluorescence for the test RNAs can give a yes-no output for binding or the relative intensity of the RNA detected may be able to give an indication of the relative binding affinity.

Table Captions

Table 1. Validated alternative unstructured linker RNA-streptavidin-binding RNA aptamer (SA_{apt}) conjugate sequences. The SA_{apt} RNA sequences consist of a core sequence (gold) (Leppek, 2014) and a variable sequence at the 5′ and 3′ ends (gold, bold and underlined) that extends a duplex that stabilises the aptamer (Leppek, 2014).

Tables

Unstructured linker RNA sequence	Streptavidin-binding RNA aptamer (SA_{apt}) RNA sequence	Reference
ACACACACACACACACACAC	**GCAUGCAU**ACCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUC GCGGGCCGGG**AUGCAUGC**	Phillips *et al*., 2018; Norouzi *et al*., 2019; Henderson *et al*., 2019
UUUUUUUUUUUUUUUUUUUU	**UGUGUG**ACCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUCGC GGGCCGGG**CACACA**	Phillips *et al*., 2018
UUUUUUUUUUUUUUUUUU	**UAGAG**ACCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUCGCG GGCCGGG**CUCUA**	Phillips *et al*., 2018
	GUGUGACCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUCGCG GGCCGGG**CACAC**	Phillips *et al*., 2018
	AUGCAUGCACCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUC GCGGGCCGGG**GCAUGCAU**	Phillips *et al*., 2018
	CGAUCGAUACCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUC GCGGGCCGGG**AUCGAUCG**	Henderson *et al*., 2019
AAUAAUAAUAAUAAUAAUAAU	**AUGCAUGC**ACCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUC GCGGGCCGGG**GCAUGCAU**	Phillips *et al*., 2018

Acknowledgements

We thank members of A.J.C.'s research group (University of Portsmouth, UK) between 2014 and 2021 for helpful discussions and technical support. We thank Dr TJ Ragan (University of Leicester, UK) for helpful discussions and critical reading of the manuscript. This work was supported by funding from the Biotechnology and Biological Sciences Research Council (BB/I532988/1 to A.J.C. and BB/L017628/1 to A.J.C.) and a University of Portsmouth bursary (to D.L.C.). Funding for the open access charge was from the Biotechnology and Biological Sciences Research Council.

Conflict of Interest Statement

A.J.C. is a named inventor on patents that relate to aspects of the work reported here. University of Portsmouth has been granted the following patents: US9777268B2 (US patent) and EP2732047B1 (European patent). C.A.H., D.L.C. and H.A.V. have no conflicts of interest to declare.

Figure 1. Generation of functional-RNA arrays by *in vitro* transcription and *in situ RNA* capture.

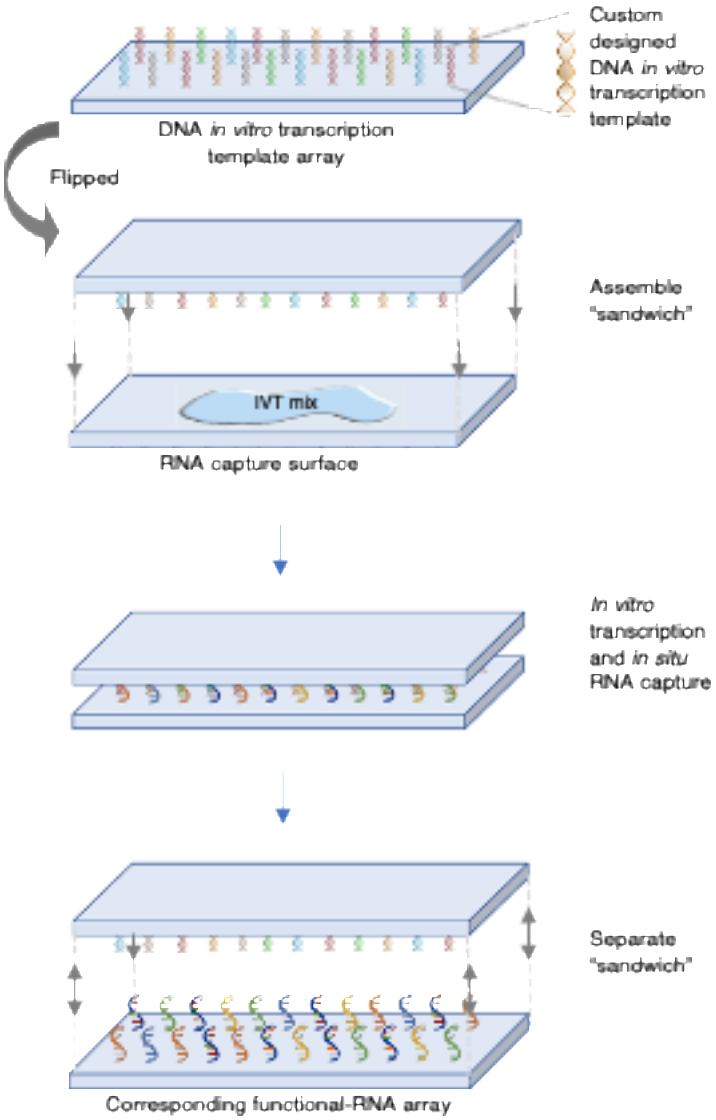


Figure 2. Design of the *in vitro* transcription template.

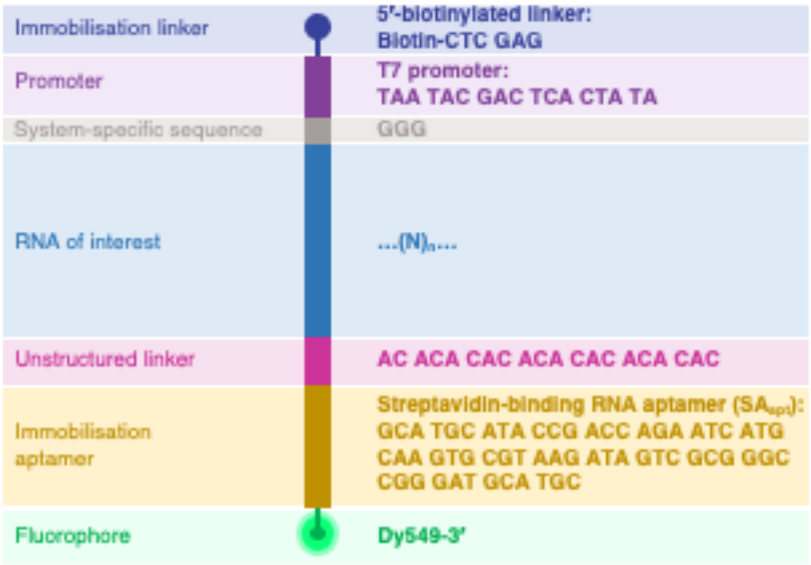


Figure 3. Predicted secondary structures for an RNA of interest (MicA) transcript.

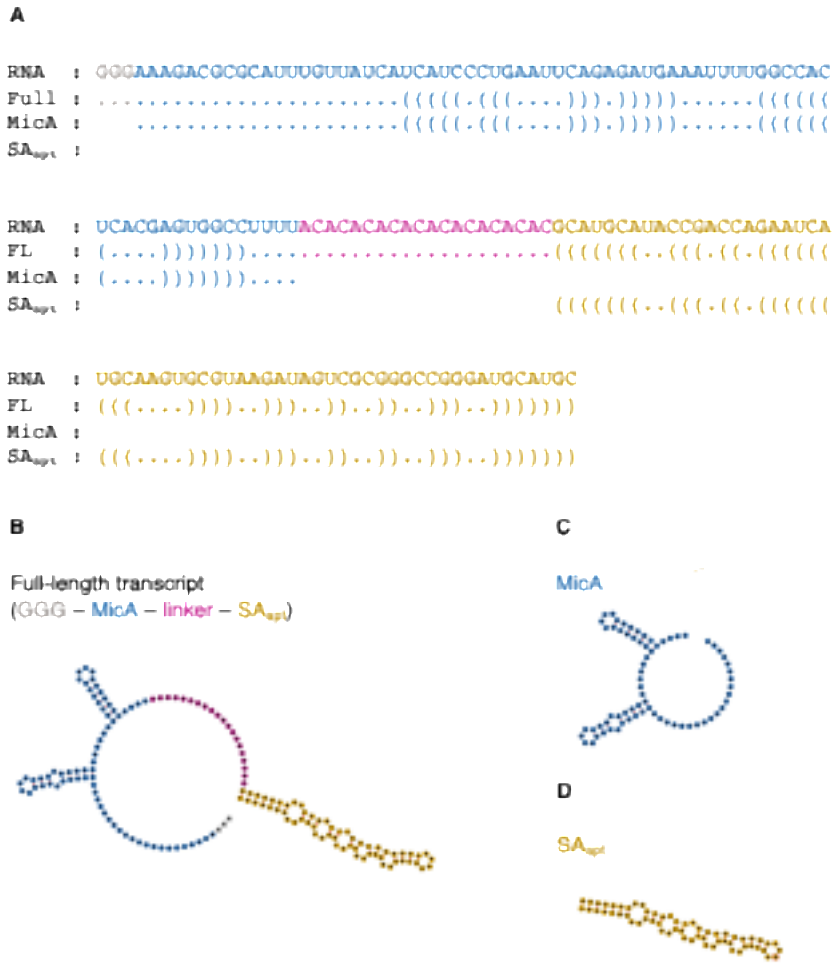


Figure 4. Spotting, visualisation and quantification of the DNA *in vitro* transcription template array.

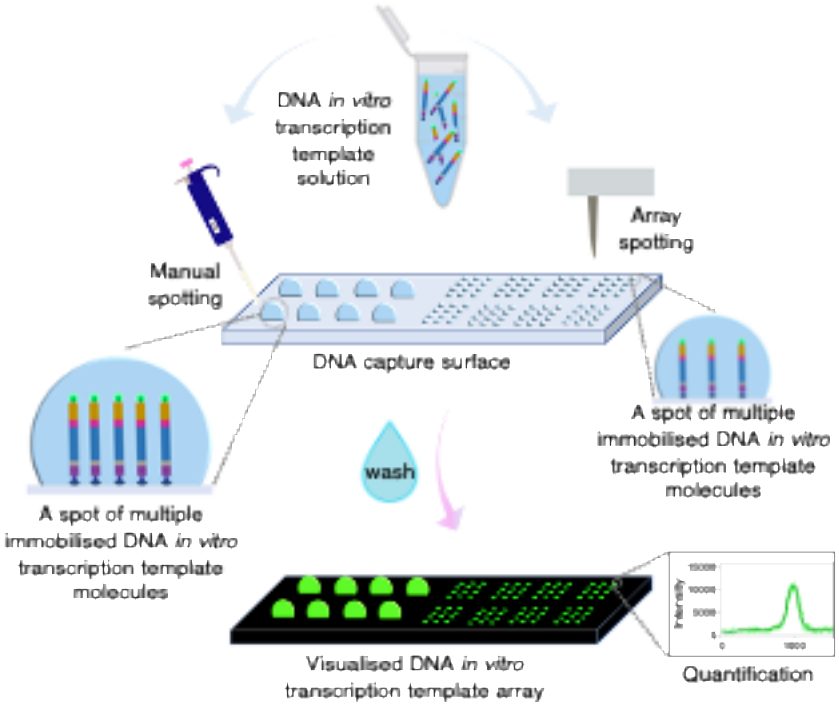


Figure 5. *In vitro* transcription and *in situ* RNA capture.

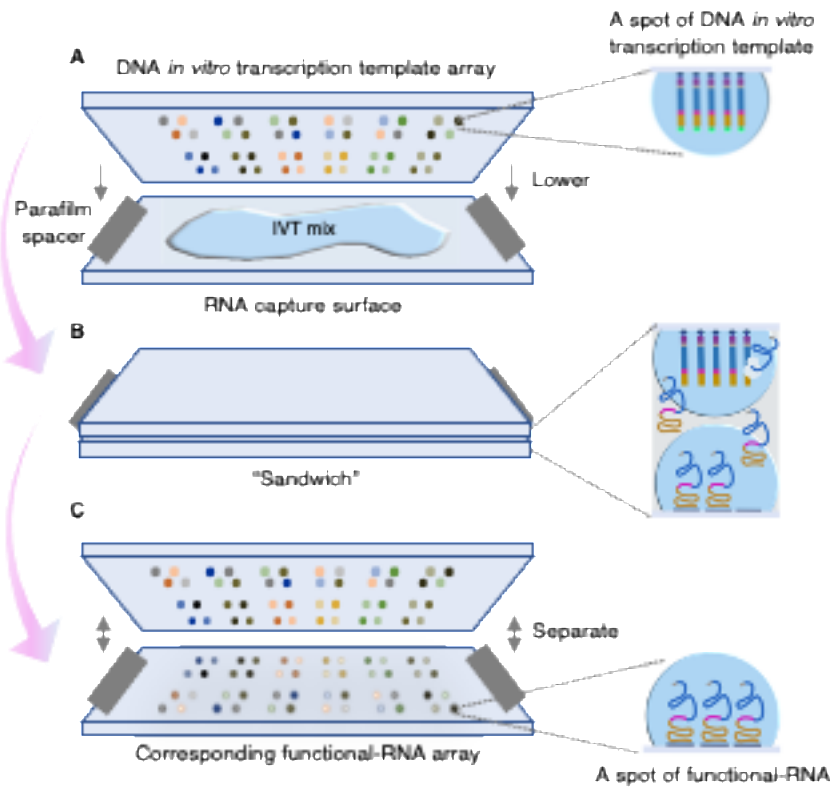


Figure 6. Probing the functional-RNA array with linker probe.

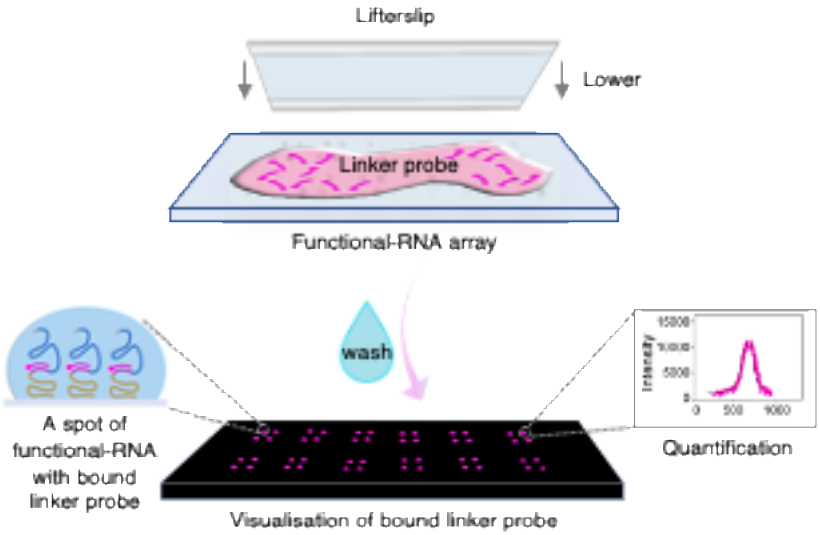
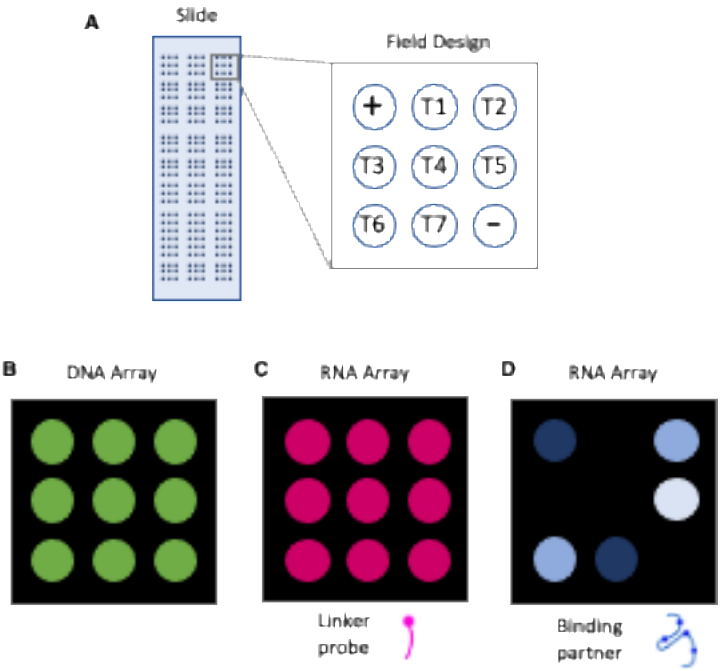


Figure 7. Application of functional-RNA arrays to the detection of RNA-RNA interactions.



UPR16 form

FORM UPR16

Research Ethics Review Checklist

Please include this completed form as an appendix to your thesis (see the Research Degrees Operational Handbook for more information

Postgraduate Research Student (PGRS) Information			**Student ID:**	654947	
PGRS Name:	Daniela Lopes Cardoso				
Department:	Biological Sciences	**First Supervisor:**	Prof. Anastasia Callaghan		
Start Date: (or progression date for Prof Doc students)	**October 2016**				
Study Mode and Route:	Part-time ☐	MPhil ☐	MD ☐		
	Full-time ☒	PhD ☒	Professional Doctorate ☐		

Title of Thesis:	**Therapeutic Targeting of the small RNA GcvB in *Actinobacillus pleuropneumoniae***
Thesis Word Count: (excluding ancillary data)	44794

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES ☒ NO ☐
b) Have all contributions to knowledge been acknowledged?	YES ☒ NO ☐
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES ☒ NO ☐
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES ☒ NO ☐
e) Does your research comply with all legal, ethical, and contractual requirements?	YES ☒ NO ☐

Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	N/A

If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:

All research carried out in the Callaghan Group, and this work specifically, involves molecular research with no ethical implications. The entire programme has been reviewed under the School of Biological Sciences ethics screening procedure (Certificate: 5802-8F52-07C7-040B-6EFB-47AF-A8A4-2E53).

Signed *(PGRS)*:		**Date:** 10/03/2021